REGULATION-INDUCED DISTORTIONS: A GEOMETRIC INTEGRATION

By

Charles William Needy

A DISSERTATION PRESENTED TO THE
GRADUATE COUNCIL OF THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1974

ACKNOWLEDGEMENTS

I am especially indebted to Milton Z. Kafoglis, chairman of my dissertation committee, for his superlative guidance and encouragement. The large influence that his theory of output maximization has had on my work will be apparent to readers familiar with his research in this area. What may not be apparent, however, is that my interest in regulation theory was sparked by his classroom lectures. Moreover, my knowledge in this field was continually reinforced during the more than two years that I served as his research assistant, since he generously chose to assign to me research projects that closely paralled my dissertation topic. As if that weren't help enough, he was also responsible for my being able to secure two grants from the Florida Public Utility Research Center for continued work on this dissertation.

I am also greatly indebted to the other members of my dissertation committee, Sanford V. Berg and Roger D. Blair, for their valuable criticism and counsel. I am particularly grateful for their skillful instruction, which heightened my interest in industrial organization, and for their strong recommendation that I select a dissertation topic early in my coursework (advice which has proven subsequently to be of inestimable value).

In addition, I received helpful comments during various

stages of my thinking from Professors Rafael Lusky, Blaine R. Roberts, and David L. Schulze, as well as from fellow student David Kaserman.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS. ii

LIST OF FIGURES vii

ABSTRACT. ix

Chapters

1 INTRODUCTION 1

2 INPUT DISTORTIONS IN THE SINGLE-PRODUCT FIRM 6

Introduction and Summary 6

Zajac's Geometric Model of the AJ Effect. 7

Simplyfying Assumptions 7
Profit Hill and Constraint Plane. 9
Constraint Curve and Isoquants. 12
Feasible Primary and Secondary Objectives . .14
Other Extensions of the AJ Effect 17

The Controversy Over Rate Base Padding 18

Zajac's Challenge 19
Westfield's Defense 22

Reconciliation of the Padding Controversy. . . .31

An Integrated Model Proposed. 32
Determinate Solution with Non-productive Capital. 37
Determinate Solution with Price Conspiracy. 43
Alternative Solution with Zajac's Procedure 45

Page

Empirical Significance of the Padding Controversy. 50

3 OUTPUT DISTORTIONS IN THE SINGLE-PRODUCT FIRM. 54

Introduction and Summary. 54

The Contractionary Output Effect. 59

Capital as a Normal Factor. 59
Capital as an Inferior Factor 60

The Expansionary Output Effect. 64

Expansion with Positive Marginal Pricing. . . 65
Expansion with Negative Marginal Pricing. . . 75

Output Effects and Their Analogs. 83

Isoquants and Indifference Curves 84
Expansion Path and Income Consumption Curve 96
Cost Employment and Price Consumption Curves. 100

AJ Distortions Versus Monopoly Restriction. . .105

Undesirable Input Effects 106
Undesirable Output Effects. 107

4 INPUT AND OUTPUT DISTORTIONS IN THE MULTI-PRODUCT FIRM. 113

Introduction and Summary. 113

Multi-Product Models of Output Distortion . . .115

Clemens Model of Price Discrimination 115
Modified Clemens Model: the Constrained Firm. 120

Multi-Product Models of Input Distortion. . . .135

Input Distortions with Identical Production Functions. 136
Input Distortions with Different, Fixed Production Functions. 139
Input Distortions with Different, Variable Production Functions 141

Page

Multi-Product Model of Input and Output Distortions, Given Variable Production Functions. 143

Optimal Combination of Input and Output Distortions. 143
Relationship Among Input and Output Distortions. 148
Alternative Managerial Preference Functions. 152

5 CONCLUSION 154

BIBLIOGRAPHY. 160

BIOGRAPHICAL SKETCH 163

LIST OF FIGURES

FIGURE Page

2.1 PROFIT HILL AND CONSTRAINT PLANE 11

2.2 LOCATION OF EFFICIENT POINT P 13

2.3 SOLUTION FOR RATE OF RETURN MAXIMIZATION 16

2.4 GEOMETRIC SOLUTION TO RATE BASE PADDING 20

2.5 INITIAL EQUILIBRIUM 24

2.6 ORIGINAL COST VALUATION 27

2.7 SOLUTION FOR STAGE III PRODUCTION 33

2.8 THE HEART SHAPED CONSTRAINT CURVE 39

2.9 JUXTAPOSITIONAL SOLUTIONS FOR RATE BASE PADDING 47

3.1 CONTRACTIONARY OUTPUT EFFECT 61

3.2 CONTRACTIONARY OUTPUT EFFECT WITH INFERIOR FACTOR 63

3.3 DERIVATION OF THE FIRM'S DEMAND CURVE 73

3.4 BREAK-EVEN SOLUTION WITH PRICE DISCRIMINATION 76

3.5 COMPARISON OF OUTPUT AND INPUT DISTORTIONS 85

3.6 COMMUNITY OUTPUT DISTORTIONS 94

3.7 COMPARISON OF FIRM AND CONSUMER EFFICIENCY LOCI 97

3.8 PRICE CONSUMPTION CURVE AND ITS ANALOG 101

3.9 EFFECT OF PRICE DISCRIMINATION ON CONSUMER AND PRODUCER SURPLUS 108

FIGURE Page

4.1 CLEMENS' MODEL OF PRICE DISCRIMINATION AND THE MULTIPLE-PRODUCT FIRM 117

4.2 STANDARDIZATION OF COST CURVES FOR THE MULTIPLE-PRODUCT FIRM 123

4.3 THE CONSTRAINED, MULTIPLE-PRODUCT FIRM 126

4.4 DERIVATION OF THE MARGINAL ASSESSMENT CURVE 128

4.5 EQUILIBRIUM SOLUTION WITH FIRST-DEGREE PRICE DISCRIMINATION AND DIFFERENT CAPITAL INTENSITIES 134

4.6 EQUILIBRIUM WHEN PRODUCTION FUNCTIONS ARE IDENTICAL 137

4.7 EQUILIBRIUM WHEN PRODUCTION FUNCTIONS ARE FIXED AND DIFFERENT 140

4.8 EQUILIBRIUM WHEN PRODUCTION FUNCTIONS ARE VARIABLE AND DIFFERENT 142

4.9 COMBINED EFFECT OF INPUT AND OUTPUT DISTORTIONS ON THE CONSTRAINED EQUILIBRIUM POSITION 144

4.10 DETERMINATE EQUILIBRIUM SOLUTION FOR THE CONSTRAINED FIRM USING BOTH OUTPUT AND INPUT DISTORTIONS 149

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida in Partial
Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

REGULATION-INDUCED DISTORTIONS: A GEOMETRIC INTEGRATION

By

Charles William Needy

June 1974

Chairman: Dr. Milton Z. Kafoglis
Major Department: Economics

This paper geometrically integrates the input and output effects induced by simple rate-of-return regulation of constrained firms. Because the output distortions have received relatively little attention in the economic literature, the paper's first objective was to demonstrate output distortions to be at least as important theoretically as the highly touted input distortions. Since each was shown to be as potentially valuable to the firm as the input distortion, it was found that the output distortions together pose a threat to social welfare that is at least as serious as that posed by the input distortion alone.

The input distortion resulted because the firm's effective cost of capital was reduced by capital's "premium," the nonproductive value of capital as a rate base and reducer of excess profits. Moreover, the constraint encouraged the firm to employ consumer cross-subsidization to enlarge the rate base by expanding output. This use of price discrimination

altered the marginal rate of substitution of the regulated goods for their unregulated substitutes (a Type 1 output distortion); and for multi-product firms it also altered the marginal rate of substitution of one regulated product for the other (a Type 2 output distortion).

It was shown that the input distortion becomes increasingly influential (relative to the size of the firm) with increases in the firm's elasticity of factor substitution. The importance of the Type 2 output distortion was shown to be determined by the amount of divergence between the capital/labor ratios of the firm's products. In addition, it was demonstrated that the effectiveness of the Type 1 output distortion is positively related to the price cross-elasticity of demand between the regulated products of the firm and their unregulated substitutes.

The three-part classification of distortions as input, Type 1 output, and Type 2 output was intended to be descriptive of the influence that each has on the productive or allocative process. In addition, a four-part classification was employed to subdivide the distortions into the AJ, Westfield, KW, and quasi-Westfield categories, which are suggestive of the causes of the inefficiencies. The AJ and Westfield effects, for example, occurred because a capital premium was created by the constraint. The KW and quasi-Westfield effects, on the other hand, resulted from consumer cross-subsidization, which in turn was encouraged by the presence of the capital premium. This four-part classificatory device satisfied the paper's second objective: to demonstrate that input and output distortions are analogous to one another. Accordingly, the AJ and Westfield effects were shown to be analogous to the KW and quasi-Westfield effects, respectively.

The recognition of this symmetrical relationship led to the construction of a determinate model (the paper's third objective), which explained the manner in which the distortions could be optimally combined to achieve constrained profit-maximization. Because a trade-off was shown to exist among the three distortion types, it was concluded that amendments to the rate-of-return constraint should dampen all three distortions simultaneously. An amendment which is narrowly designed to reduce only one inefficiency may well exacerbate the other two. Since regulation was demonstrated to be theoretically more costly than the recognition of input distortions alone would imply, it was concluded that economists should place less emphasis on regulating market performance and more on changing market structure into a more competitive environment. Where regulation is still deemed desirable, it was felt that greater attention should be paid to rate structures to ensure that output inefficiencies receive at least as much attention as the input distortions with which they are combined.

CHAPTER 1

INTRODUCTION

The economic literature that had been developed out of the original Averch-Johnson article concerning regulation-induced distortions has been largely preoccupied with input distortions created by the simple rate-of-return constraint. Except for the pioneering work done by Kafoglis and Wellisz, little has been written about output distortions. The little that has been said has only pointed out that input distortions raise average costs that in turn pull the firm's output below the social optimum (i.e., where P = MC). It comes as quite a surprise, then, to learn that Averch and Johnson (A and J) consider output distortions to be equally important as (and not simply incidental to) the input distortions they identify.

Chapter 2 of this work briefly reviews the AJ effect and shows how Westfield carries the AJ input analysis to its logical extreme. In essence, Westfield proves informally that the constrained firm will have a profit incentive to purchase useless capital when its production process is physically saturated with the capital input. Chapter 3 identifies two AJ output effects, one being contractionary and the other expansionary. The former receives little attention because it is incidental to the AJ input effect and, as such, has already received much attention in the literature. The latter effect, however, has been examined rigorously only by Kafoglis and Wellisz (K and W), so it is of primary concern. This expansionary output effect is produced when the simple rate-of-return constraint induces the firm to sell a portion of its output at below-MC prices. Unless product demand is perfectly inelastic, the lowering of the marginal price will increase the quantity of product demanded. In this way a firm having excessive profits can expand its rate base and trim its earnings at the same time, since the expanded output justifies--and helps to finance--an increase in capital stock. In addition, this output distortion (of the marginal rates of substitution of the regulated products for unregulated products) may lead to an inefficient use of total energy resources. The most dramatic possibility is that low-income classes may subsidize certain sectors. Thus, both distributional equity and economic efficiency may be frustrated simultaneously.

In reference to the behavior of a restrained profit-maximizer, A and J explain the efficacy of below-MC pricing but not the extent to which it might be profitably employed. With reference to the restrained output-maximizer, Kafoglis extends the concept of below-MC pricing to its "probable" extreme. He uses a first-degree price discrimination model to demonstrate that regulation can induce output expansion far beyond the Pareto-optimal level. Even though Kafoglis uses an output- rather than profit-maximizing objective, the important implications of his analysis can be almost fully applied to the AJ profit-maximizer as well. Therefore, it is apparent that the restrained profit-maximizer may also be induced to expand output beyond the social optimum by the rate-of-return constraint. This application of Kafoglis' analysis in Chapter 3 is highly significant, considering that A and J discuss below-MC pricing only in terms of third-degree price discrimination and that they fail to identify the potential, total output effect with respect to the social optimum. Moreover, Wellisz examines below-MC pricing only within the framework of peakload pricing, and he too fails to identify the potential, total output effect.

Even though Kafoglis carries the below-MC pricing concept to its "probable" extreme, he does not carry it to its logical--and, perhaps, improbable--extreme. He does not consider the possibility that negative marginal pricing could be the profitable result of the combination of price discrimination and regulation. Apparently, this possibility has been overlooked in the rest of the economic literature as well. The analysis in Chapter 3 shows negative marginal pricing to be a workable policy when the firm reaches the point of market saturation before exhausting excess earnings and when the regulatory agency prevents the firm from acquiring capital in inefficiently large proportions. This result is not unlike the Westfield input effect, which becomes viable when either market- or factor-saturation is reached before excess earnings are eliminated. Because of this analogous relationship between the two influences, negative marginal pricing might be conveniently referred to as the quasi-Westfield output effect. This nomenclature emphasizes the symmetrical relationship between input and output distortions while acknowledging that the Westfield analysis does not actually encompass output distortions.

Two of the prime objectives of this work are accomplished in Chapters 2 and 3. These goals are (1) to elevate the expansionary output distortion to the status which it deserves--a status at least equal to that now enjoyed by the input effect--and (2) to define and emphasize the analogous relationship that exists between input and output distortions. Accordingly, the AJ and Westfield effects will be shown to be analogous to the KW and quasi-Westfield effects, respectively.

The third major objective, to demonstrate how the firm might optimally combine the distortions to achieve constrained profit-maximization, is accomplished in Chapter 4. This fourth chapter not only produces a determinate equilibrium solution that combines both input and output distortions, it yields this solution within the framework of a multi-product firm. The theoretical analysis of regulation-induced inefficiencies has too long been confined to a study of single-product firms, so this extension by itself is quite significant.

By employing a multi-product model of the constrained firm, a third type of distortion will be identified that is to be labelled a "Type 2" output distortion. Until now, it has not been adequately distinguished from either the "Type 1" output distortion or the input inefficiency. Finally, implications of these theoretical developments with regard to rate design will be examined in the conclusion of the thesis, Chapter 5.

CHAPTER 2

INPUT DISTORTIONS INDUCED BY CONSTRAINT

Introduction and Summary

The exposition of formal theoretical models is initiated in this chapter, which demonstrates that effective rate-of-return constraints may lead to input distortions in the production process. Because the interpretation of the Averch-Johnson (AJ) effect has great bearing on subsequent chapters, Zajac's [1970] treatment of the AJ model is carefully reviewed at the outset of this chapter.[1] This graphical model reaches the now-familiar Averch-Johnson conclusion that when the market price of financial capital is less than the rate of return permitted by the regulatory agency, the constraint[2] encourages the profit-maximizing firm to use more capital relative to labor than is needed to minimize costs at the output level it selects. This conclusion is now popularly referred to as the "AJ effect."[3] The latter part of this chapter examines Westfield's argument that the constrained, profit-maximizing firm may purchase useless or idle capital under output conditions normally encountered by decreasing cost industries.

---

[1] The Averch-Johnson article employs only a little geometry, relying heavily on nonlinear programming and the Kuhn-Tucker theorem. The theorem is discussed by Bailey [pp. 20-32].

[2] Because this work is solely concerned with the fair rate-of-return constraint, all subsequent usages of "constraint" are intended to refer to this usage only, unless otherwise stated.

Zajac's Geometric Model of the AJ Effect

Simplifying Assumptions

The Zajac [1970] model is based upon the following assumptions: a single firm uses two variable factors, capital K and labor L, to produce a quantity q of a single good which can be sold at price p.[4] Other endogenous variables are determined by the inverse demand function,

$$p = p(q); \tag{1}$$

the total revenue function,

$$R = qp; \tag{2}$$

---

[3] An exception is Kahn's reference to the input distortion as the "AJW effect" which gives equal recognition to S.H. Wellisz, who developed a similar model almost simultaneously with (but independent of) Averch and Johnson. See Kahn [1971, pp. 49-59]. The Wellisz article emphasizes the importance of regulation-induced output distortions (between peak and off-peak markets) while the AJ paper rigorously examines input distortions.

[4] See Averch and Johnson [pp. 1054-56] for the assumptions as originally presented. The symbols used here are borrowed from Zajac since his graphical model is a major reference in this work.

the production function,

$$q = q(K,L); \qquad (3)$$

and the profit function,

$$\pi = pq - (iK + wL), \qquad (4)$$

where i is the cost of capital and w is the wage rate of labor. From the first three equations it is apparent that total revenue can be expressed as a function of the inputs K and L, so that

$$R = R(K,L). \qquad (5)$$

Given that both factors are variable, the model must be considered a long-run analysis either in the sense that sufficient time is allowed for the completion of all desired input adjustments or in the sense that the firm is designing its plant from scratch. The latter of these two interpretations seems more sensible, because the static model represents reality at an instant in time. The former interpretation is feasible, of course, since all functions could be assumed to remain constant in the long-run.

Regulation (as envisioned in this model)[5] affects the firm's decisions in only one regard: it imposes a constraint which requires that the rate of return to capital not exceed a "fair" rate of return f. The model, in short, assumes rate-of-return regulation rather than detailed rate regulation as is predicted by many commissions. To avoid trivial solutions, it is assumed that f is greater than i (so the firm will survive in the long-run) but below the attainable, profit-maximizing rate of return (so the firm has an effective constraint with which to contend).[6] The regulatory constraint can be stated as

---

[5] In reality, regulation goes far beyond this simple restraint.

$$\frac{pq - wL}{K} \leq f, \qquad (6)$$

when depreciation and regulatory lag are assumed nonexistent. Zajac rearranges (6), using equation (4), to obtain

$$\pi \leq (f - i)K. \qquad (7)$$

If the constraint is effective, (7) can be written as an equation which holds at the point where constrained profits are a maximum. In addition, constraint (7) and all other functions are assumed to be twice differentiable, while the necessary second-order conditions are assumed to be fulfilled. Also, factor costs and isoquants are given. Finally, technological change and uncertainty are dismissed for simplicity, so we have a purely static model.

Profit Hill and Constraint Plane

From equations (1) and (5) it is clear that, given w and i, p, q and π depend solely on K and L. The profit function specified by equation (4) can therefore be shown in Figure 2.1 as a profit hill using π, K and L coordinates. The above assumptions concerning differentiation and concavity conditions assure that only one value of π exists for each K,L combination; that the profit hill is single-peaked; and that only one point, the peak, is tangent to the K,L horizontal plane. When interpreted as an equality, constraint (7) forces the firm to operate below the hill's peak on a constraint plane shown hinged on the labor axis. The degree to which the plane swings away from the K,L horizontal plane is determined by equation (7). Since the constraint plane is assumed to be effective, it cuts through the profit hill, delineating a prohibited "bump" on top of the hill where the firm may not operate. Because the constraint plane is hinged on the labor axis, the firm maximizes profit when operating along the intersection of the profit hill and the constraint plane at the point of maximum K. The curve which results from the intersection of the hill and plane is called the "constraint curve." The fact that rate-of-return regulation permits operation at any point on the constraint curve is itself powerful evidence that a wide variety of outcomes are possible under this simple type of restraint. About all that this constraint can be expected to accomplish is "profit" control.

---

[6] Moreover, the solution is trivial when f=i in the sense that a secondary objective would be needed to make the solution determinate. See Averch and Johnson [p. 1055] and Baumol and Klevorick [p. 174]. In reality, regulatory lag could create a situation where f=i or f<i (at least in the short-run).

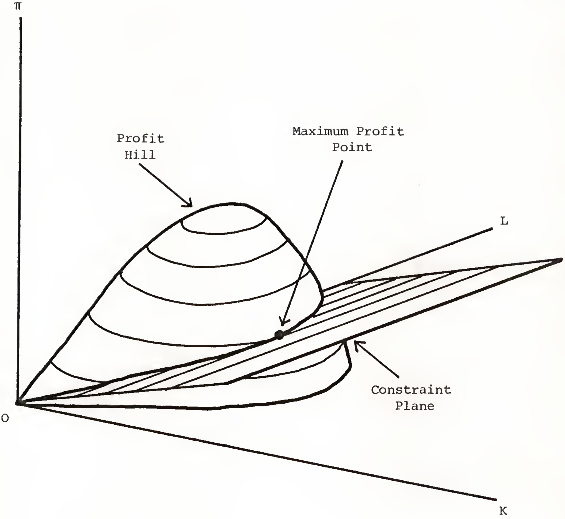


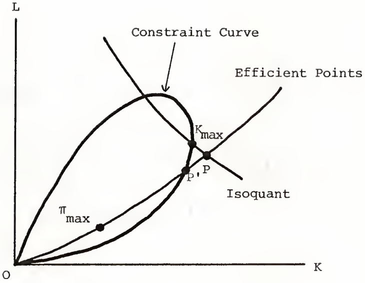
FIGURE 2.1

PROFIT HILL AND CONSTRAINT PLANE

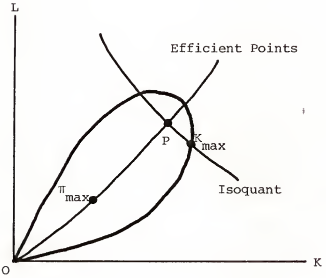
Constraint Curve and Isoquants

In Figure 2.2 the constraint curve is projected down onto the horizontal K,L plane. The firm is not allowed to operate in the area where the prohibited bump is circumscribed by the constraint curve. Since factor costs and the isoquants are given, a locus of efficient points can be drawn which would be the expansion path of an unconstrained firm. From the inverse demand function, $p = p(q)$, it is known that the isoquants are also isorevenue curves. From equation (4) it follows that profit is maximized along any isoquant at the efficient point, where input costs are minimized. The locus of efficient points, therefore, must pass through the π_{max} point at the peak of the profit hill. The firm operates at K_{max}, at which point it wastefully utilizes too much K relative to L.

To prove that the firm uses too much K for the chosen output, Zajac shows it to be impossible that the locus of efficient points would intersect the isoquant passing through K_{max} outside the constraint curve. As shown in Figure 2.2(a) this outcome would require efficient point P to be more profitable than K_{max}. And P', also located on the locus of efficient points but closer to π_{max}, must be more profitable than P. This would mean, however, that P' is more profitable than K_{max}, which is inconsistent with the fact that profit is maximized along the constraint curve at K_{max}. Consequently, the locus of efficient points must lie above K_{max} and intersect



(a) Impossible Location of P



(b) Only Possible Location of P

FIGURE 2.2

LOCATION OF EFFICIENT POINT P

the isoquant through K_{max} at a point inside the constraint curve as shown in Figure 2.2(b). Therefore, the firm produces the K_{max} level of output employing too much K relative to L.

Feasible Primary and Secondary Objectives

It is interesting to note that as the regulator becomes increasingly lax, allowing the constraint plane to swing upward, the forbidden "bump" on the profit hill diminishes in size. As a result, K_{max} in Figure 2.2(b) moves closer to π_{max} as the forbidden area shrinks. Conversely, as f is forced closer to i the constraint plane swings downward, approaching the horizontal base of the profit hill. This enlarges the prohibited area, forcing K_{max} farther away from π_{max} and farther away from the efficient point on the isoquant through K_{max}. Therefore, as the constraint becomes increasingly more restrictive (i.e., as f approaches i) the input distortion grows increasingly larger.[7] This conclusion is not too surprising, since it is obvious that no input inefficiency will exist when f is so great as to be ineffective. It should be equally clear that when f = i the profitless firm cannot pursue its sole objective, profit-maximization. In this unlikely case the firm would be indifferent to operation at any two points on the horizontal, constraint plane. A determinate solution would result only when a second objective (i.e., sales or output maximization) is assumed to be present. As stated earlier, Zajac's basic model avoids these two extremes by assuming f to be less than the π_{max} rate of return and greater than the cost of capital.[8]

---

[7] This conclusion is more fully explained by Baumol and Klevorick [pp. 174-75] and Scherer [p. 533]. In addition, a geometric proof is offered by Westfield [pp. 429-32] and a rigorous mathematical proof is produced by Takayama [pp. 255-60].

Even when f does exceed i, the firm may still choose some maximand other than profit as its objective. For example, if the firm is solely intent upon maximizing rate of return, Zajac shows that it would be indifferent as to its position on the constraint curve as shown in Figure 2.3. The rate of return is everywhere the same along the constraint curve. The equilibrium solution becomes determinate only when the firm is assumed to have a secondary objective, which in this example could be cost-minimization. In this case the firm would operate at position A, because this is the only point on the constraint curve that is intersected by the locus of efficient points. Alternatively, if the firm's primary and secondary objectives are rate of return and output maximization, respectively, the firm will meet both objectives at position B. Only at point B is the constraint curve tangent to an isoquant curve (and both curves in turn tangent to a level line of fK + wL). Because f is greater than i, the level line of fK + wL must be steeper than that

---

[8] Averch and Johnson [pp. 1055-56] themselves consider these two extreme cases, which are also given cursory examination by Zajac [p. 118].

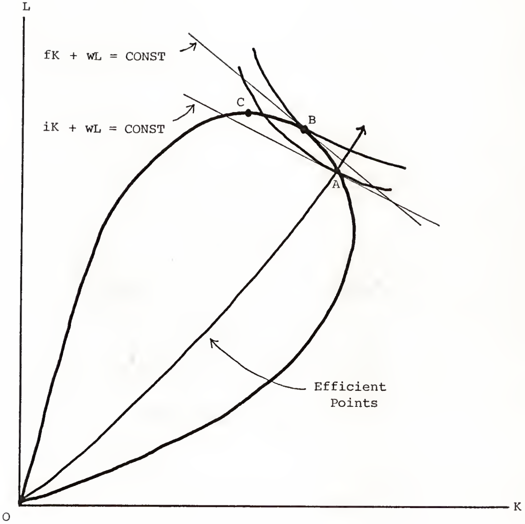


FIGURE 2.3

SOLUTION FOR RATE OF RETURN MAXIMIZATION of iK + wL, which intersects point A. This means that point B is necessarily northwest of point A and is located on the constraint curve between points A and C,[9] since the level line of fK + wL is assumed to be negatively sloped. At position B, then, the firm will undercapitalize--not overcapitalize--because it is operating above the locus of efficient points. This reversal of the AJ effect also occurs if the firm's secondary objective is sales (i.e., revenue) maximization, and the firm would again reach an equilibrium position at B.[10] It is clear that rates of return are equal at points A and B while costs are greater at point B. The higher costs at B are offset by higher revenues.

Other Extensions of the AJ Effect

The graphical exposition is now complete for Zajac's model, which will become the basis for the extended geometry in Chapter 4. The Zajac model is only a small part of the growing body of literature that has been inspired by the original Averch-Johnson article. Bailey and Coleman [1971], for example, explore the effects of regulatory lag (assumed nonexistant in the Zajac model) on the AJ effect. In addition, Baumol and Klevorick [pp. 179-80] disprove the claim that the constrained, profit-maximizing firm will necessarily use a higher capital-labor ratio than that employed by the unconstrained firm. Although these and other insights are not sufficiently basic to this paper as to warrant exposition here, one allegation is highly relevant to the theoretical results in the next three chapters. This item is the claim that the constrained firm will increase allowable profits by purchasing nonproductive capital with which to pad its rate base. The remainder of this chapter will be devoted to an analysis of this important proposition.

---

[9] Point C could only be tangent to a line having zero slope.

[10] The reversal of the AJ effect was first demonstrated by Bailey and Malone [pp. 129-42].

The Controversy Over Rate Base Padding

Zajac labels the claim that the firm may want to pad its rate base with useless capital an "ancient allegation" and handles it summarily at the end of his article [1970, pp. 124-25]. He takes the opposing view that the firm would not willingly purchase useless capital when it could buy productive capital for the same price. For the sake of simplicity and economy of exposition, Zajac attempts to disprove the allegation within the geometric framework of a model already introduced (Figure 2.1 above). Westfield's arguments in favor of the proposition will follow a summary of Zajac's reasoning against it.[11] Because no previous attempt has been made to integrate both arguments within the same model, the Zajac and Westfield presentations will be combined geometrically in an effort to reconcile the opposing views.

---

[11] It might seem more appropriate that a rigorous defense of the allegation precede--not follow--the well-reasoned argument against it. Such would be the precedure of the affirmative and negative participants in a formal, verbal debate. However, since supporters of the old view are

Zajac's Challenge

In an attempt to finally settle the issue of rate base padding, Zajac lowers the profit hill in Figure 2.1 so that it looks like that in Figure 2.4. The added cost of redundant capital K* causes the entire hill to sink lower by an amount iK*. The profit function becomes

$$\pi = pq - wL - i(K + K^*). \qquad (8)$$

Similarly, the constraint can now be restated as

$$\pi \leq (f - i)(K + K^*), \qquad (9)$$

which means that the constraint plane will rise to a new position that is parallel to the old. The new, higher plane intersects the K-axis at -K*. Even though the rise in the constraint plane has a positive effect on allowable profits, this is more than offset by the negative effect imposed

---

reportedly guilty of a misconception, this summary of the dispute will follow legal procedures in placing Zajac on the stand first. It should then be apparent why Westfield, the most eloquent defender of the old belief, is being brought to trial. More to the point, the Westfield defense is quite structurally dissimilar to the Zajac model. The defense must be accompanied by new analysis and terms. On the other hand, Zajac's prosecuting remarks can be discussed in the familiar terms that have only recently been introduced.

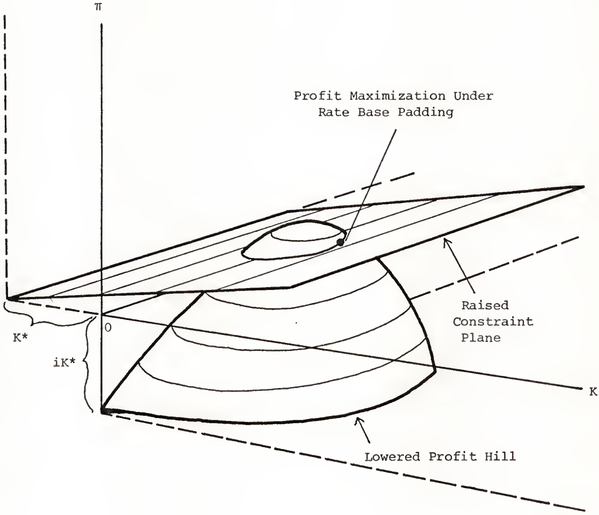


FIGURE 2.4

GEOMETRIC SOLUTION TO RATE BASE PADDING

by the fall of the profit hill.

Zajac's result is reasoned as follows. If, with the purchase of useless capital, the firm were maximizing profits it could substitute productive capital for the useless capital K*. Because K and K* cost the same, K* can be traded for an equal amount of productive capital. This situation means that the firm can now remain on the same isoquant while cutting back on labor employment. Since the rate base is unchanged, the reduction in wL pushes profits up to a prohibited level. The firm could react by purchasing more productive capital and using it to replace labor. This moves the firm outward along the same isoquant to the constraint curve. When it reaches the constraint curve, the firm can further increase profits by moving along that curve to position K_{max}, shown in Figure 2.1. In this way Zajac arrives at the conclusion that

> . . . the constrained profit maximum of [Figure 2.1], using only productive capital, is always higher than the constrained useless-capital profit maximum of [Figure 2.4]. [1970, p. 125]

Nevertheless, Zajac's argument that the firm will never wish to purchase useless capital is correct only within the framework of his formal model, which assumes--by implication rather than specification--that the firm will never produce an output so great that capital is used beyond the point of physical saturation. Because he ignores the possibility that the constrained firm might operate outside Knightian production stage II,[12] Zajac fails to come to grips with the theoretical arguments advanced by Westfield.

Westfield's Defense

Writing several years in advance of the Zajac article and using a graphical model that is quite unlike Zajac's Figure 2.2(a), Westfield shows that profits of the constrained, profit-maximizing firm will not necessarily decrease--and indeed may increase--when the capital supplier raises the price of capital. This result would explain, Westfield argues, why suspected conspiracies between electrical utilities and their capital supplies could be profitable to both parties.

Westfield's geometric model uses two basic graphs: one static and the other comparative-static. The former is based entirely on assumptions already introduced above. The comparative-static graph, however, needs an added assumption that the firm will discount future earnings with the rate of interest r. A distinction must be made, therefore, between the average acquisition cost of capital i and the interest cost r of holding capital stock. In addition, both graphs require the introduction of a few new terms. The first of these is "Total Net Revenue Product of Capital," a Robinson concept[13] which is represented in Figure 2.5 by curve T_k. Using equations (4) and (5), the T_k function can be written

---

[12] As will become apparent below, the situation envisioned by Westfield is Knight's Stage III, in which the variable factor is combined with the fixed factor in wastefully large amounts. For his classic presentation, see Knight [p. 100].

$$T_k = \max_{L} \, [R(K,L) - wL] = T(K). \qquad (10)$$

This function requires that L be optimally adjusted for every K value to ensure that the marginal revenue product of labor just equals the marginal expenditure on labor (also equal to w since labor is supplied at a constant price). Restated, this requirement says that

$$\frac{\partial R}{\partial L} = w. \qquad (11)$$

Note that equation (11) is not an added assumption, since it is a logical consequence of action taken by the efficient profit-maximizer in both the Zajac and Westfield models. Moreover, even though Westfield's firm may choose to purchase K at prices above the competitive level (as will be shown below), it is not necessary to drop the earlier assumption that K is bought in a competitive market. One of the competitive suppliers may be more than willing to enter a price-raising conspiracy with its unusual customer. Note also that K is purchased beyond the point at which its marginal revenue product and marginal cost are equal. Excessive capital acquisitions are needed to reduce total profit

---

[13] Westfield [p. 430] credits the terminology to Robinson. Note that T_k is net of labor costs but gross with respect to capital costs.

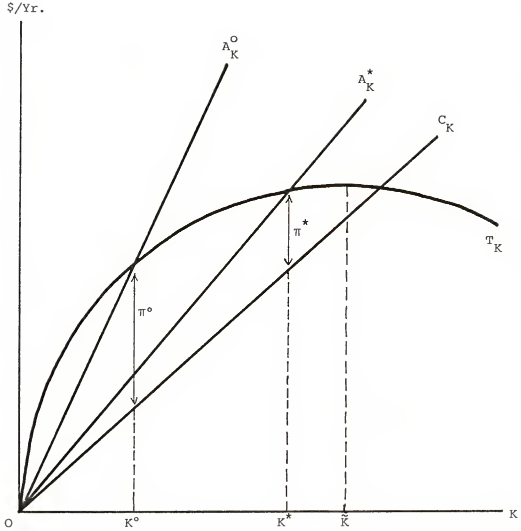


FIGURE 2.5

INITIAL EQUILIBRIUM

to the level at which the firm earns only the fair rate of return f. This is the familiar AJ effect, demonstrated earlier.

Figure 2.5 bears a close resemblance to a vertical plane in Zajac's model. That is, the ordinate measures earnings (and capital costs) and the abscissa measures physical units of K. Another similarity is the ray through the origin labeled A^*_k, which shows the "Maximum Allowable Total Net Revenue Product" of capital. In the Zajac model such a line might rather be called the "constraint line"--distinguishing it from the constraint curve--since it is a line drawn parallel to the π,K plane on the constraint plane.

Having no counterpart in the Zajac model, ray C_k in Figure 2.5 represents the total annual cost of employing various quantities of K. Its function can be defined as

$$C_k = r \cdot i \, K. \tag{12}$$

At each level of K, then, maximum total profit can be measured as the vertical distance between C_k and T_k. This profit maximum is greatest at K°, at which level the slopes of C_k and T_k are equal. The slope of T_k is $dT(K)/dK$ and the slope of C_k is ri. Because f is assumed to exceed i but fall short of the profit-maximizing rate of return, the constrained firm will be in short-run equilibrium when

$$A^\circ_k > A^*_k = T_k > C_k.$$

The firm maximizes profits π^* by employing a K^* level of capital.

To demonstrate that the restrained, profit-maximizing firm may be indifferent to--or benefit from--increases in the cost of capital, Westfield employs comparative-statics in a second graph. The equilibrium solution of Figure 2.5 is reproduced in Figure 2.6, so that T_k again intersects A_k above K^*. The model is thrown into disequilibrium by a sudden growth in demand, which increases p and causes T_k to immediately shift upwards to one of the three new positions shown.[14] The firm will purchase new capital goods to supplement the old capital stock K^* so that output can rise to meet the increased demand. The total capital cost curve C_k remains unchanged if the incremental capital, labeled I, can be obtained at the initial price i. Likewise, the constraining ray A_k remains unchanged as long as the regulator continues to value the new capital by the same criteria that he used to value the old capital. Assuming that the original T_k curve shifts upward to the new T_k (rising) position, equilibrium will be re-established at $\bar{a}$ as a result of new investment I equal to $K_1 - K^*$. At the same time, profits increase from distance c^*a^* to distance $\overline{ca}$.

At this point there are only three possible new equilibrium solutions, depending upon the choice of the three

---

[14] Westfield points out that the same increase in T_k could also result from a decrease in tax rates or wage rates.

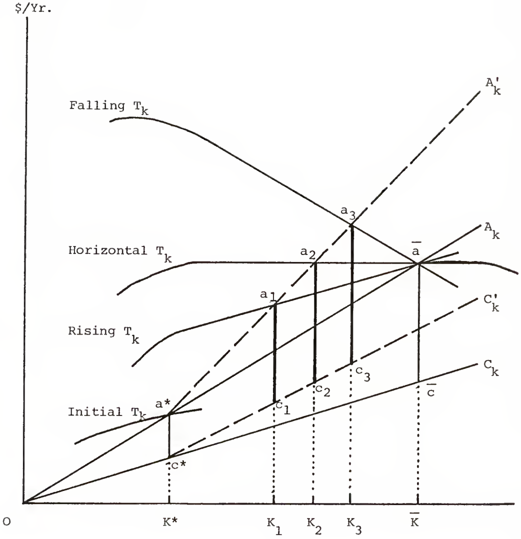


FIGURE 2.6

ORIGINAL COST VALUATION

alternative T_k positions (i.e., rising, horizontal or falling). When the possibility that capital costs i may rise to a higher incremental cost i' is recognized, the number of feasible solutions increases to six. If i' exceeds i, the C_k and A_k rays will each be increased by the same percent, because the allowable rate of return is figured as a percent of capital stock to begin with. Since the initial valuation of K* is unchanged, ray C_k kinks at c* and swings upward by a given percent to C_k'. Similarly, ray A_k kinks at a* and swings upward by an equal percent to A_k'. The increased total cost of capital function becomes

$$C_k' = r\ i\ K^* + r\ i'\ I. \qquad (13)$$

It is assumed that the regulator expands the rate base by the amount of new investment, so that at the various levels of capital beyond K* ray C_k' is effectively the rate base.

Even though the slopes of A_k and C_k are increased by equal proportions, the allowable net revenue grows by a larger, absolute amount at each higher level of K than does the total cost of capital because A_k is initially steeper than C_k. In other words, the increase in total capital costs resulting from the capital price rise is more than offset by the corresponding increase in the allowable, total net revenue product of capital. Consequently, a potential exists for the regulated firm to benefit from a price conspiracy with its capital suppliers. Whether the firm actually loses, breaks-even, or benefits from an increase in i is entirely determined by the slope of the T_k curve in the neighborhood of the four constraint and capital cost curves.

Accordingly, there are three separate cases which Westfield considers. If it is first assumed that T_k rises in the relevant region, the new equilibrium intersection of T_k with A_k will occur at $\overline{K}$ and the firm will earn profits equal to distance $\overline{ca}$, as earlier explained. If i' is greater than i, however, T_k intersects with A_k' at K_1 and the firm's profit is only c_1a_1, which is necessarily smaller than $\overline{ca}$. As long as T_k rises, then, the firm is unable to take advantage of the potential increases in allowable profit that are generated by the conspiracy. Furthermore, profits are reduced by higher cost of capital.

In the second case, T_k is horizontal when it intersects the two alternative constraint curves, A_k and A_k'. A price rise for new investment would cause the equilibrium level of capital to decrease from $\overline{K}$ to K_2, but profits would remain unchanged since $\overline{ca}$ must necessarily be equal to c_2a_2. In such a situation the firm would be indifferent to a rise in the price of capital goods.

The last possibility is that T_k is falling as it intersects A_k and A_k'. Here the firm will react to higher capital costs by reducing its factor purchases from $\overline{K}$ to K_3, but profit is shown to increase from $\overline{ca}$ to c_3a_3. This third situation is the most interesting of the three, because it is here demonstrated that the constrained profit-maximizing firm actually favors an escalation of capital prices. And what amounts to the same thing, the regulated firm may tend to pad its rate base with nonproductive capital. The T_k curve peaks and falls due to either market saturation or physical saturation. That is, T_k falls either because the marginal revenue product or the marginal physical product becomes negative. Westfield admits that the former situation is most improbable since it requires that the firm operate along the inelastic portion of its demand curve. The rate of return constraint alone could not force the firm to expand production beyond the point at which marginal revenue is zero. On the other hand, physical saturation will surely occur because capital must eventually reach its intensive margin regardless of the given level of labor input.

At capital's intensive margin $MPP_k = 0$ and T_k reaches its peak. Curve T_k does not fall from this peak until MPP_k becomes negative. Westfield attempts to satisfy this negativity requirement by arguing that excessive additions to capital stock may ". . . contaminate the production process . . . . like redundant rain . . ." [p. 440] so that output maintenance would require additional units of labor. Such contamination might take the form of storage costs or the tendency of unemployed equipment to get in the way. The situation can be worsened, then, if the regulator insists that all capital stock be actively employed, because this would make it all the more likely that the production process will be corrupted. Having demonstrated that the T_k curve may fall, Westfield is able to conclude that the constrained, profit-maximizing firm may favor price conspiracies or acquisition of useless capital stock.

Reconciliation of the Padding Controversy

Three reasons come to mind for combining the Zajac and Westfield arguments within the basic Zajac model [i.e., Figure 2.2(b)]. First, even though Zajac's conclusion is incongruent with the Westfield conclusion that precedes it, Zajac does not directly respond to the Westfield arguments. This is unfortunate, since the Westfield analysis can be handled economically within the framework of the model introduced in the Zajac article. The integration of the opposing arguments here will produce just the kind of juxtaposition needed and which Zajac seems to have avoided. Second, Westfield is forced to abandon his geometry when explaining the nature of physical saturation. His graphs are inadequate for this purpose because they lack axes for both factor inputs. On the other hand, isoquants in the Zajac model solve this problem easily. The production of a geometric, whole model out of geometric, verbal, and mathematical parts should make the Westfield analysis more quickly comprehensible and better integrated. Third, the Zajac model is well known while the Westfield graphs have been slow to catch fire in the literature. Students of regulation theory may more quickly comprehend Westfield's reasoning when it is translated into familiar terms.

An Integrated Model Proposed

The integrated model shown in Figure 2.7 is logically consistent with the Zajac analysis and dependent only upon the familiar Zajac assumptions.[15] The negative and positive sloping isoquant Q can be superimposed onto Zajac's Figure 2.2(b), because his use of a negative sloping isoquant in that figure does not preclude the possibility that the isoquants are also positively sloped in some undefined regions. A great increase in product demand, for example, could expand Zajac's constraint curve so far outward that it would encompass the positive portions of some isoquants (like Q and Q' in Figure 2.7). Therefore, the isoquants and constraint curve are compatible with Zajac's analysis.

At the same time, point A in Figure 2.7 satisfies the Westfield requirement that there be some finite quantity of capital $\tilde{K}$ which is capable of producing output Q when combined with $\tilde{L}$. Output Q cannot be produced unless at least $\tilde{L}$ amount of labor is employed in the production process.

---

[15] Unlike Zajac, Averch and Johnson [p. 1054] specifically assume that the marginal physical products are always positive for both factors. Because this distinction has an important bearing on the present analysis, the assumptions are here attributed to Zajac instead of Averch and Johnson.

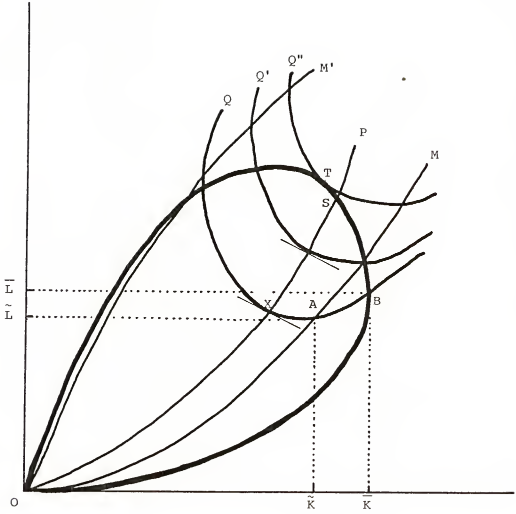


FIGURE 2.7

SOLUTION FOR STAGE III PRODUCTION

Point A in Figure 2.7 corresponds to the peak of curve T_k in Westfield's Figure 2.6. In the two graphs it can be observed that isoquant Q reaches its minima and T_k reaches its maxima above the same level of capital $\tilde{K}$. In cooperation with $\tilde{L}$, $\tilde{K}$ is the smallest quantity of capital which just barely produces Q. If L is fixed at $\tilde{L}$, the marginal product of capital is zero at $\tilde{K}$. The marginal rate of technical substitution of labor for capital (i.e., the slope of Q at point A) is also zero, because it equals the ratio of the MPP_K to MPP_L. The value of this ratio is zero, of course, since the numerator is zero.

Ridge line M intersects the minima of an infinite series of isoquants. Everywhere along this isocline the slope of the intersecting isoquant, the MRTS of L for K, and the MPP of K are all equal to zero. Ridge line M is the intensive margin of capital, so the firm's production stage III lies to the right of this line. State II is bounded by ridge lines M and M', the latter being the extensive margin of capital. Isocline P is the locus of points along which each isoquant is tangent to an isocost curve. This locus would be the firm's expansion path if it were unconstrained. Everywhere along isocline P the ratio of input prices is equal to the MRTS of L for K.

The equilibrium intersection of the constraint curve with an isoquant occurs at position B, at which point the production process uses capital in inefficiently large proportions. The input distortion caused by the constraint can be separated into two perverse influences. The first is the AJ effect, which induces the firm to move from efficient point X up to capital's intensive margin at point A. The second influence is the "Westfield effect," which encourages the firm to move beyond point A to position B in a region considered uneconomic by unconstrained firms.[16] At point B the restrained firm equates the MRTS of L for K (now negative) with the ratio of its private input-prices (this ratio also being negative). The private cost of capital is less than the market cost because capital serves as the rate base. The AJ effect pulls the private cost of capital below the market cost to a level near zero.[17] The Westfield effect causes private cost of capital to become negative.

---

[16] A good case could be made for calling the entire distance XB the AJ effect, since the AJ analysis could produce equilibrium at point B if only one AJ assumption is relaxed (i.e., the stipulation that $MPP_K > 0$). On the other hand, the separation of the input distortion into the two effects identified here is descriptively useful. Furthermore, the separation should seem less arbitrary in Chapter 3, where the AJ and Westfield input distortions are shown to be analogous to two types of output distortion.

In addition, the view that the AJ effect is different from the effect of useless capital purchases is subscribed to by Zajac [1971, p. 118] and McNicol [pp. 433-34]. Finally, Johnson [1973, p. 91] himself concludes that "the so-called Averch-Johnson effect of overcapitalization does not as a general case involve 'gold-plating' or purchase of plant solely to be held idle . . . ."

[17] Because of the AJ assumption that MPP of both inputs is always positive, the AJ effect can force the private cost of capital so low that it asymptotically approaches zero. For this reason the AJ effect does not actually move the firm from point X to position A--but, rather, to a position infinitely close to position A.

All isoquants greater than Q' intersect the constraint curve to the left of ridge line M. This observation means that the revenue-maximizer whose secondary goal is cost-minimization will operate at position S, where the locus of efficient points intersects the constraint curve. If, instead, this firm's secondary goal is output or sales maximization, it will operate at position T and produce output Q". In any event, the revenue-maximizer having either of these secondary objectives will operate at outputs greater than Q' and thus be free of the Westfield effect. This conclusion has general validity and should not be interpreted as characterizing only the case shown in Figure 2.7. Point S, for example, can never be located on the positively sloped portion of an isoquant, because the isoquant must be tangent to a negatively sloped isocost curve at point S (by definition). This ensures that point T will not fall on the rising portion of an isoquant since the isoquant's slope at T is even steeper in a negative direction than it is at S. Point T will always be located to the left of point S as long as the fair rate of return f exceeds the market cost of capital i. Only the constrained profit-maximizer, then, can be influenced by the Westfield effect.

In the Westfield analysis the firm conspires to raise the price of K so that it can retreat from point B to a position closer to A on isoquant Q. Even though the firm operates somewhere between points A and B, and even though the market price of K rises, the isoquant's tangent at point

B continues to measure the firm's effective (private) input-price ratio. Distance AB, then, is a consistent measure of the Westfield effect even though the firm's actual operating position lies closer to capital's intensive margin at point A.

When the firm withdraws its operations from point B toward A, there is no compelling reason to believe that the firm will finally relocate somewhere midway between points A and B, as shown in the Westfield analysis. It is more logical to conclude that the price of $\tilde{K}$ will be driven up so high that operations will be fully retrenched to position A. At this point only $\tilde{K}$ physical units are employed while inflated K prices raise the value of the rate base. In terms of Westfield's Figure 2.6 this would mean that A_k' rises to an intersection with T_k at its peak instead of at point a_3. It might be argued that the firm will not raise capital prices enough to fully retreat to point A, because risk of conspiracy detection increases along with capital price increases. However, since the analysis is theoretical anyway, it is just as reasonable to assume that the conspiracy is fully successful as it is to suppose it to be only partially successful. Therefore, it is the stronger of these two cases which shall be analyzed below.

Determinate Solution with Nonproductive Capital

It has been stated that the firm will benefit from a price conspiracy that moves its operating position from B to A on isoquant Q (the strong case for the Westfield effect). This result can be more easily illustrated by using the numerical example in Figure 2.8. It will first be shown that the firm will benefit by purchasing useless capital, and then this result will be shown to be analogous to a price increase of capital. The isoquant Q in Figure 2.8 is constructed on the assumption that the K axis measures only productive capital. Nevertheless, this axis will also be used to measure purchases of useless K*, since the firm can buy K* at the same price as K.

A numerical solution

If the firm is initially producing output Q at position A, it employs 180 units of L and 300 units of productive capital K. The firm cannot remain at A because the fair rate of return f is violated at any position within the constraint curve. The firm must expand its rate base if it is to continue to produce the highly profitable output Q. This is accomplished if 60 units of K are purchased, which moves the firm to the allowable position at B. Unfortunately, the added units of K contaminate the production process, so the firm must simultaneously purchase an additional 30 units of L to maintain output at Q. Total revenue is the same at B as at A, but the added units of K and L serve to lower the rate of return to an allowable level.

It must be remembered, however, that the profit-maximizer would be foolish to operate at position B. Point B serves

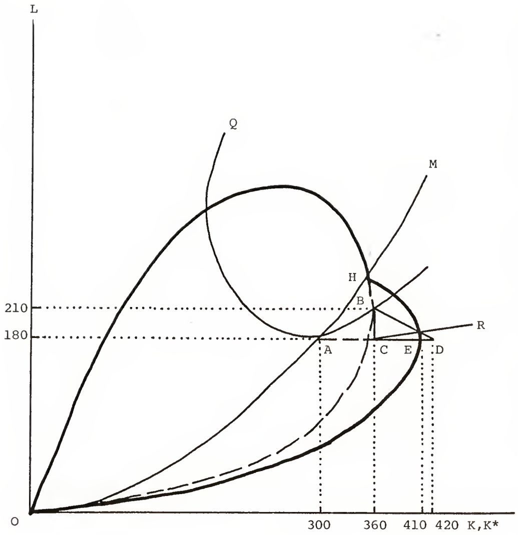


FIGURE 2.8

THE HEART SHAPED CONSTRAINT CURVE

only as the firm's alternative position. It is a marker that can be used to determine the firm's effective input-price ratio. Instead of buying 60 units of K, the knowledgeable firm will purchase 60 units of useless K*. In this way the firm expands its rate base to 360 units of K and K*, while continuing to employ only 300 units of K in the production process. The firm's actual isoquant is the concave curve QAB, but the firm's effective isoquant is the curve QACED. This latter curve assumes that productive capital K is employed only up to the point of its intensive margin and that all further additions of capital are K*. Consequently, the firm's operating position is A and its effective position is now C.

At position C total revenue, capital costs, and the rate base are the same as they were at B, but 30 units less labor is now purchased. Because profits were at the permitted limit at B, profits must exceed that limit at C by the value of the unemployed labor. If the firm dissipates excess profits by rehiring the 30 labor units it would return to position B--hardly a wise choice. But given a K/L price ratio of 1/2, the firm can exhaust the excess earnings on capital additions alone by acquiring 60 more units of useless K*. That is, 30 L units can be exchanged for 60 K* units, moving the effective position from C to D. Actually, the firm will not have to forfeit all of the excess profits because the rate base will be expanding along with the expansion of K* purchases.

The firm's equilibrium position is easily determined. The firm has 60 K* units of wealth at its disposal and a portion will be spent on purchases of useless capital and a portion retained for profit increases. Symbolically this becomes

$$\Delta K^* + \Delta\pi = 60. \tag{14}$$

Equation (7) can be restated as

$$\pi = (f-i)(K + K^*), \tag{15}$$

and if (f-i) = 1/5, then

$$\Delta\pi = \frac{K^*}{5} . \tag{16}$$

Combining equations (14) and (16) produces

$$\frac{6\Delta K^*}{5} = 60, \tag{17}$$

so that $\Delta K^* = 50$. This means that the firm expands its rate base by 50 K* units to 410 units of capital. The expanded rate base justifies a profit increase equal in value to 10 K* units. In effect, the firm has exchanged 25 of the 30 L units for 50 K* units and sold the remaining 5 L units for a profit increase.

A geometric solution

The graphical solution is equally simple. Isocost line BD limits one side of triangle BCD. Ray R through the origin at point C has a slope of (f-i), which equals 1/5 in this example. As the firm substitutes K* for L, it moves rightward along isocost line BD. The southeasterly movement along line BD is accompanied by a corresponding northeasterly movement along ray R, which reflects the steady increase in allowable profit level (as the rate base is expanded along line CED). The equilibrium intersection of constraint ray R with line BD occurs directly above point E. The constraint ray R should not be confused with Westfield's constraint line or Zajac's constraint plane. Unlike the other two concepts, constraint ray R measures the value of allowed profit increases in terms of labor units--not profit or revenue units. For this reason it lies on the K,L plane--not above it.

Profit hill.--The firm's movement of effective positions from A out to E will have an effect on the profit hill (and thus the constraint curve) that has not yet been discussed. Like the "productive" isoquant curve QAB, the profit hill is initially drawn under the assumption that only productive capital will be purchased. The acquisition of useless capital, however, causes the "effective" isoquant QACED (and all other isoquants) to be perfectly elastic beyond capital's intensive margin. The new profit hill must be constructed using an infinite series of effective--not productive--isoquants for its derivation. Because the effective and productive isoquants are identical to the left of ridge line M, the profit hill outside the stage III region is not affected by the firm's clandestine purchasing policy. To the right of ridge line M, however, the profit hill must be higher at all points.

Heart shaped constraint curve.--The new constraint curve can be approximated without use of a laboriously constructed, new profit hill. It has already been shown, for example, that the firm's rate of return at position C violates the constraint. Position C, therefore, must lie inside the new constraint curve. At the other extreme, it is already known that the firm's rate of return at position D is below the allowable level f. Point D, then, necessarily lies outside the new constraint curve. By similar reasoning it follows that the new constraint curve must intersect point E, since the firm earns a fair rate of return at E. It can be deduced in this manner that the new constraint curve will be heart shaped,[18] intersecting points O, H and E. This must be one of the few instances in microeconomic theory where a firm is guided to a profit-maximizing solution by its heart--as well as its head.

Determinate Solution with Price Conspiracy

The curves in Figures 2.7 and 2.8 might best be viewed as planning curves, since the model is static and both factors are fully variable. Even though it has been analytically useful to consider the firm's rate of return at several positions in Figure 2.8, the firm itself is assumed to proceed immediately to equilibrium position E. It should be evident that the firm can reach position E either through acquisitions of useless capital, as has just been explained, or through price conspiracy with one of its competitive suppliers of capital.

---

[18] Zajac himself recognizes that the constraint curve could be ". . . heart shaped, with the point of the heart at its origin, or kidney shaped" [1970, p. 119]. In the present analysis, the constraint curve will be heart shaped in the presence of rate base padding only if it is assumed to be tear shaped in the absence of rate base padding.

Using the former method, the firm expands its rate base from a productive 300 units to an inefficient 410 units. The addition of 110 nonproductive units of K* wastefully inflates the rate base by 37 percent. Such a large acquisition of useless capital stock could be very costly to store and difficult to explain to the regulatory agent. For these reasons it might take the form of a very expensive, highly technical machine (i.e., a computer) that takes up little space and is apparently useful. On the other hand, the firm may find it less risky to enter a price conspiracy in which capital prices are inflated 37 percent. In this way the firm could purchase 300 K units for the price of 410 K units. Like the useless capital ploy, the conspiracy enables the firm to fully extricate its plant operations from the "uneconomic" third stage of production while at the same time helping it to expand its rate base.

Depending upon the point in time at which the conspiracy begins, the price-fixing effects on the model can take many geometric forms. If the price agreement is formalized only after the firm has reached some operating point inside stage II near ridge line M, the modified constraint curve could look something like the heart shape in Figure 2.8. This example would correspond closely to Westfield's comparative-static example in Figure 2.6. However, it is not likely that the firm would wait until it had purchased nearly 300 K units before agreeing to pay higher capital prices. Such delay would force the firm to pay questionably high prices on the few incremental units of K.

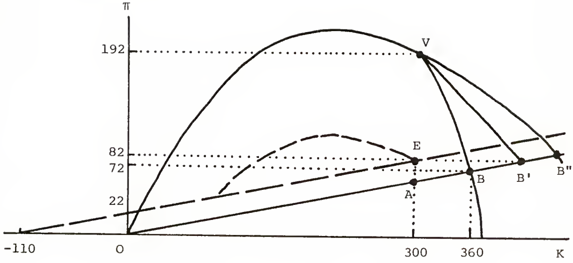
It seems more reasonable, then, that the firm would bargain for higher prices at the outset of its operations, so that the inflation of K unit costs can be held down to the less noticeable 37 percent increase. Because this maneuver makes K* purchases unnecessary and undesirable, the constraint curve retains the tear shape of Figure 2.7 instead of acquiring the heart shape of Figure 2.8. Because each K unit increases in nominal value by 37 percent, the slope of the constraint plant increases by 37 percent. At the same time, the profit hill is diminished at each level of K by 37 percent of the value (at initial prices) of K employed. The resulting constraint curve lies inside that shown in Figure 2.7 and exhibits its K_{max} position at point A.

Alternative Solution with Zajac's Procedure

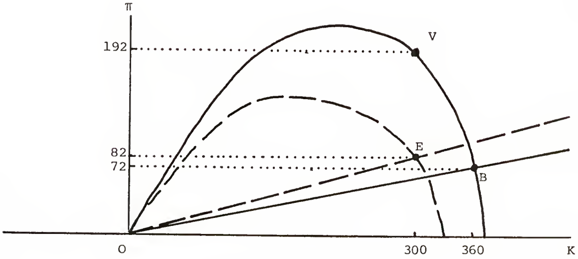
The model being used here for the contrast of the Zajac and Westfield results differs from the Zajac model in only one important regard (other than the use of positively sloped isoquants): both K and K* are measured on the abscissa. Because of this slight modification, a determinate solution can be resolved on the two-dimensional K,L plane even when the firm acquires K* units. In the unmodified Zajac model, however, the introduction of inoperative K* purchases creates expositional problems. Zajac is forced out of his simple two-dimensional model into a cumbersome, verbal comparison of two three-dimensional models. Even though Zajac's approach to the rate base padding issue is more difficult to follow, it will be used here to demonstrate that the results obtained above are valid regardless of the theoretical apparatus employed. However, instead of comparing two three-dimensional models, the present analysis uses a simple two-dimensional graph that measures only K--not K and K*--on the abscissa. In this way the figure will be consistent with Zajac's model while simpler in form.

Westfield results with useless capital

Accordingly, the Zajac approach to the issue is taken in Figure 2.9(a). The abscissa measures only physical units of productive capital, and the ordinate measures profit in units that each have a value equal to one unit of capital. Consequently, the 20 percent slope of constraint line OBB'B" means that F exceeds i by 20 percent of K. The figure represents a vertical slice through Zajac's Figure 2.4, so the negatively sloped portion of total profit curve OVB is a locus of all feasible K_{max} points. The intersection of the



(a) Effect of Nonproductive Capital



(b) Effect of Price Conspiracy

FIGURE 2.9

JUXTAPOSITIONAL SOLUTIONS FOR RATE BASE PADDING profit and constraint curves occurs at position B, at which point the firm is earning an acceptable rate of return (points A, B and E here correspond to the identically labeled points in Figure 2.8). The constrained profit-maximizer will not remain at position B. This firm sells 60 of its 360 K units, thus freeing 30 L units[19] which it also sells. This refund is equivalent to the value of 120 K* units, but the firm elects to purchase only 110 K* units which inflates its cost of 300 K units by 37 percent. As a result, the constraint line rises to the position of the dashed line by an amount equal to $(f-i)\Delta(K + K^*) = \Delta K^*/5 = 22$ units. At the same time, the profit hill--and therefore the profit line--falls by $i\Delta K^*$. Since profit units are valued as being equivalent to K units, the profit line falls a full 110 units to the position of the dashed curve. Equilibrium is established at point E, where the firm's profit is 10 units greater than its alternative position at point B.

Zajac results with useless capital

When Zajac manipulates the profit hill and constraint curve in this same manner, he arrives at the opposite conclusion that profit falls--not rises. This apparent inconsistency

---

[19] It is here apparent that Figure 2.9(a) has its disadvantages, too, since the labor input is not shown. When compared with Figure 2.9(a) and Zajac's Figure 2.4, the modified Zajac model in Figure 2.8 has the two advantages of showing the effect of K* purchases while (1) also showing both inputs and (2) avoiding the disconcerting comparison of three-dimensional graphs.

is explicable. In the example used here, the firm's MPP_k is negative for K levels exceeding 300 K. In Figure 2.9(a) this means that profit line OVB will have a (negative) slope in excess of unity throughout the relevant segment VB. However, if instead that segment has a slope equal to unity (i.e., if $MPP_k = 0$ in the 300 K to 360 K region), profit curve OVB' will produce the same 82 profit units at position E or B'. In this case (described in Westfield terms as a horizontal T_k curve), the firm is indifferent to K* purchases that do not exceed 110 units. Finally, the profit curve OVB" describes Zajac's example in which the VB" segment possesses a slope less than unity (because $MPP_k > 0$ in the relevant region). It necessarily follows that profit is less at point E than at point B" and that the firm will avoid K* purchases.[20]

Westfield results with price conspiracy

It has already been shown that a price increase can yield results equivalent to the purchase of useless capital, but so far the effects of a price increase on the profit hill have only been described verbally. However, these effects can now be conveniently illustrated geometrically, since Figure 2.9 lends itself well to a demonstration of price conspiracy. In Figure 2.9(b), then, the firm is assumed to find price conspiracy more attractive than additions of useless K* stock. If the firm were myopic, it might purchase 299 K units at the regular price and then surreptitiously negotiate a 1100 percent price increase on the last K unit purchased. It seems more realistic to suppose, however, that the firm will negotiate a less noticeable 37 percent price increase on all 300 K units that it intends to buy. If the firm is initially constrained by ray OB, the 37 percent price increase on capital units increases the slope of OB by 37 percent, swinging it upward to the dashed line OE. At the same time, the profit curve falls by 37 percent of the value of capital employed at each level of K. At a level of 300 K, for example, total profit will fall by .37(300) = 110 units, since profit is measured in units that are equal in value to K units. The firm's equilibrium position at E in Figure 2.9(b) is identical to its equilibrium position in Figure 2.9(a), so factor price conspiracy is shown to be equivalent to the purchase of useless capital stock.

---

[20] Zajac's Figure 2.4 [1970, p. 125] is technically incorrect, because the profit hill is drawn with such a steep slope that it would yield a profit curve at least as steep as OVB in Figure 2.9(a). Since both axes are ungraduated, Zajac may intend that the unit value of the profit axis far exceeds the unit value of the capital axis. However, the distance K* on the abscissa is shown to slightly exceed distance iK* on the ordinate, and this is exactly the opposite of what is required. For the profit hill to be correct as shown, distance K* would have to be at least one-third shorter than distance iK*.

Consequently, Zajac's conclusion that the constrained profit maximum in Figure 2.4 will always be greater than that in Figure 2.1 [1970, p. 119] could never be valid. This small technical error is easily avoided in Figure 2.9(a) since both profit maxima are shown on the same graph.

Empirical Significance of the Padding Controversy

From the point of view of its author the Westfield effect is considered to be not only a possible outcome--but a probable outcome--of constrained profit-maximization. To this effect, Westfield concludes that

> It is possible even to contrive special cases in which our results will be "unlikely." However, we show that without first setting up a straw man, our propositions are logical consequences of profit-maximization. [p. 442]

Westfield's argument that the T_k curve (Figure 2.6) will inevitably level off is very persuasive. His two explanations [p. 440] for the existence of a negatively sloped portion of the T_k curve--if not convincing--are logically defensible. However, his work is purely theoretical and offers no compelling evidence that the T_k curve will descend (if at all) within a range of capital employment that is relevant to actual, regulated firms. It is difficult to understand, therefore, way Westfield regards his own results as representative of reality, while he regards contrary results as contrived "special cases."

If the Westfield effect is to be discovered empirically,[21] it must be found in industries which exhibit increasing returns to scale, since these industries are prone to regulation. If capital's intensive margin M in Figure 2.8 has an increasingly steep slope, scale economies will expand the constraint curve outward and move the firm's effective position E further toward (or into) production stage III. In this case the Westfield effect should be positively related to scale economies and a high degree of price elasticity (since this would also enlarge the constraint curve). Conversely, if the slope of capital's intensive margin M increases at a decreasing rate, the margin M will fall farther and farther away from the locus of efficient points P with increases in output. This situation means that the likelihood of physical saturation occurring grows increasingly smaller with increases in output. This situation describes firms which become more capital intensive (and less labor intensive) with increases in output. The existence of scale economies in this case could easily expand output to high levels at which capital's saturation point is far removed from the firm's actual operating position. Growth in the telephone industry, for example, may encourage the replacement of local telephone operators with specialized equipment. This situation would make the Westfield effect negatively related to scale economies and high levels of price elasticity. In summary, if regulated firms are truly characterized by increasing returns to scale and a positive relationship between output and capital intensity, the Westfield effect is empirically--not theoretically--problematical. That is to say, it would be least likely found in the only industries where it can possibly exist.

---

[21] In addition to Zajac, Bailey [pp. 192-95], Emery [pp. 329-36], Corey [pp. 358-73], and Kendrick [pp. 136-67] consider the Westfield hypothesis empirically unfounded. Kahn [1971, pp. 53-54] takes no position in his discussion of the Westfield proposition. Through casual empiricism it can be noted that regulated firms are not in the forefront in adopting polution abatement equipment.

Nevertheless, the Westfield proposition cannot be summarily dismissed on theoretical grounds (as Zajac attempts to do),[22] even if empiricism reveals it to be a stranger in its own home.

---

[22] Following the completion of this analysis, Professor Blair has brought to my attention a more recent article by Zajac [1972] in which he changes his original position, concluding that rate base padding indeed may be profitable to the regulated firm. Zajac demonstrates his new position with a graph that is very similar to my Figure 2.9(a) above. As was demonstrated in Figure 2.9, rate base padding through acquisition of useless capital or price conspiracy is effectively the same thing. It is surprising, then, that the recent note by Zajac [1972] makes no reference to the Westfield article.

Regardless of Zajac's change of mind, confusion about the efficacy of rate base padding continues to surface in the regulatory literature. In the autumn of 1973, for example, McNicol [1973, pp. 433-34n.] contends that the AJ effect can be proved only if (among other things) it is assumed that ". . . the isoquants are strictly convex to the origin." Figure 2.7 above proves otherwise, though I have chosen to label the positively sloped (concave) portion of the isoquants the "Westfield effect."

CHAPTER 3

OUTPUT DISTORTIONS IN THE SINGLE-PRODUCT FIRM

Introduction and Summary

The economic literature that has developed out of the original Averch-Johnson article seems to be preoccupied with input distortions created by the simple, rate-of-return constraint. With the exception of Bailey, Bailey and White, Kafoglis, Kahn [1968], and Wein, the general consensus has little to say about output distortions, except to point out that input distortions will raise average costs to the firm and reduce output below the social optimum. This conclusion is reached by Zajac, Bailey and Malone, Baumol and Klevorick, Westfield, Sheshinski, Klevorick [1966, 1971, 1973], Wichers, and Takayama. As an example, Klevorick [p. 729] writes that many papers in the literature fostered by the AJ work ". . . elevate input efficiency from the status of a means to an end to the status of an end in itself." It is also interesting to note that Baumol and Klevorick's synthesis of this same body of regulatory literature is titled "Input [emphasis added] Choices and Rate-Of-Return Regulation: An Overview of the Discussion." Moreover, within this review [p. 168] the authors reach the accepted conclusion that the AJ effect will induce input distortions that in turn pull output below the socially optimum level.

It comes as quite a surprise, then, to find that Averch and Johnson consider output distortions to be equally important as (and not simple incidental to) the input distortions identified in their original paper.[1] Placing both distortions on an equal footing, A and J begin their article by writing that "regulatory bias" has two effects--not just one.[2] These are

> (1) The firm does not equate marginal rates of factor substitution to the ratio of factor costs; therefore, the firm operates inefficiently in the sense that (social) cost is not minimized at the output it selects.
>
> (2) The firm has an incentive to expand into other regulated markets, even if it operates at a (long-run) loss in these markets; therefore, it may drive out other firms, or discourage their entry into these other markets, even though the competing firms may be lower-cost producers. [p. 1052]

The first of these is the AJ effect on input choices. The corresponding output influence is contractionary (when compared to the social optimum where P = MC) and is only incidental to (i.e., a result of) the highly touted input distortion. The second is the little-recognized AJ effect on output decisions. Here the perverting influence will expand output instead of contracting it.

---

[1] Even more surprising, A and J devote a full two-thirds of their article [pp. 1057-67] to a discussion of output distortions and only a third of the article to the input model. The latter, of course, is examined far more rigorously, which no doubt explains why "AJ effect" is normally used in reference to the input distortion alone.

[2] Wein refers to the AJ effects (1) and (2) as propositions A and B. Contrary to its cursory treatment in the literature, Wein's discussion of the second effect places it above the first in order of importance. Unlike the focus of this paper, however, Wein's approach to the subject is institutional rather than theoretical. See Wein [pp. 42-43].

Below-MC pricing is the cause of the expansionary output effect. The AJ analysis of this effect lacks analytic rigor, being largely institutional in approach. On the other hand, Kafoglis and Wellisz explore below-MC pricing (and the resulting expansionary output effect) within formal theoretical models. Consequently, the expansionary output effect is here identified with Kafoglis and Wellisz in the same way that the literature has identified the input effect with Averch and Johnson. Hence it will be convenient to refer to the effect on output of below-MC pricing as simply the "KW effect" (so named after Kafoglis and Wellisz). Like the AJ (input) effect, the KW (output) effect produces a net loss of social welfare. The allegation[3] that below-MC pricing can be more beneficial than harmful, since it may offset monopolistic restriction of output, will be shown to be a spurious claim.

Chapter 2 identified the AJ input effect and showed how the Westfield effect carries the AJ input analysis to its logical extreme. The present chapter identifies the output-distorting counterparts to these two input-distorting influences. The first counterpart is the KW effect, which will be the principal concern of this chapter because it has not yet been given a comprehensive examination in the literature. The AJ contractionary output effect will receive little attention here, because it is incidental to the AJ input effect and, as such, has already received much attention in the literature. Basically, the contractionary output effect is produced by cost increases attendant to the AJ input distortion. The KW effect, however, is an output distortion caused by below-MC pricing, which can occur independently of input distortions.

---

[3] This argument is advanced by Kahn [1969, p. 172 and 1971, p. 101].

The second counterpart is the output distortion which results from the firm's use of negative marginal pricing. The possibility of negative product prices (i.e., bribes) for favored customers or customer classes has been overlooked by both Kafoglis and Wellisz. Moreover, it has been overlooked in the rest of the literature as well. In the same way that the Westfield effect carries the AJ effect to its logical extreme, the analysis of negative marginal pricing carries the KW effect to its logical extreme.

Although the first objective of this chapter is exploration of the KW effect, the secondary objective is identification of the analogous relationship between input and output distortions. The KW (output) effect is shown to be analogous to the AJ (input) effect, and the influence of negative marginal pricing is likewise shown to be analogous to the Westfield effect. Because of this symmetrical relationship, the effect of negative marginal pricing is expeditiously referred to here as the "quasi-Westfield effect." This nomenclature calls attention to the analogous relationship between the two influences, while at the same time acknowledging that the Westfield analysis does not actually encompass output distortions.

During the last decade, economists have discussed input distortions at length. Because a clearly defined symmetry exists between input and output distortions (that are induced by simple rate-of-return regulation), it is hoped that much of the extensive input analysis can quickly be applied to output distortions. As an example, the dampening effect that regulatory lag has been shown to have on input distortions[4] could also be shown to have a like effect on output distortions. Moreover, the mitigating effect that a graduated rate of return could have on the AJ and Westfield effects[5] would be equally offsetting to the KW and quasi-Westfield effects. In addition, a tightening of the rate-of-return constraint has the same aggravating influence on output distortions as it does on input distortions.[6] Nonetheless, the transfer of these analyses from input to output space--as rewarding a venture as it might be--is outside the scope of this thesis. Instead, the primary purpose of this chapter is to demonstrate the output effects to be at least as important as the highly touted input effects.

---

[4] The influence of lagged regulation on the AJ effect was first explored by Bailey and Coleman [1970]. This influence is more recently examined by Klevorick [1973].

[5] Actually, the graduated rate of return has been shown to have the same effect on input distortions as a higher (non-graduated) rate-of-return ceiling. The graduated rate was proposed by Klevorick [1966] and successfully refuted by Wichers. Klevorick [1971] later rescinds his arguments that once supported the proposal. For further discussion of the debate, see Bailey and Coleman [1970], Kahn [1968] and Sheshinski [p. 178 n.].

The Contractionary Output Effect

The Averch-Johnson input effect has been widely discussed in the literature. Because this perverse effect on factor combinations indirectly contracts the firm's level of output, the contractionary output effect is incidental to the input effect. Consequently, the contractionary output effect has also received a lot of attention in the literature, since a discussion of the one effect would be incomplete without a discussion of the other. The now-familiar relationship between these two effects can be briefly summarized.

Capital as a Normal Factor

Averch and Johnson [pp. 1056-57] argue that the input distortion caused by the rate-of-return constraint will have a concomitant effect on output. Using a graph similar to

---

[6] The aggravating effect on output distortions is demonstrated by Wellisz [p. 36]. The worsening effect on input distortions is discussed by Westfield [pp. 429-32], Takayama [pp. 255-60], Baumol and Klevorick [pp. 174-75], and Scherer [p. 533].

Figure 3.1,[7] which shows the revenue and cost situation facing a regulated monopoly, they reason that the input distortion will increase average costs from AC_1 to AC_2. The production function is here assumed to combine two normal factors K and L in variable proportions. Regulation is claimed to have the beneficial effect of expanding production from the unconstrained monopoly level at A to the constrained output level at B. The socially desirable output level is C' if the firm is simply required to break-even (i.e., to equate AC and AR). However, the optimal level is C, which can be achieved if the firm is subsidized or allowed use of price discrimination. For simplicity, output level C' is assumed to be the regulatory objective in this case. Output C' is not reached, however, because the inefficient combinations of factors raises AC_1 to AC_2. The reduction in output from C' to B, then, is the contractionary output effect.

Capital as an Inferior Factor

Westfield [p. 430] was the first to point out that the

---

[7] Figure 3.1 differs from the Averch-Johnson graph [p. 1057] in only one respect: the firm's allowable rate of return is here assumed to be included in the average cost curve. This assumption causes the firm's break-even level of output to be indicated by the intersection of the AC and AR curves, rather than by some point to the left of that intersection (the approach used by A and J). Because the graphs in this chapter are quite self-contained, this assumption is used to simplify exposition. In the rest of this paper, however, graphs using AC are sometimes discussed in parallel with the Zajac model, so this assumption is not made.

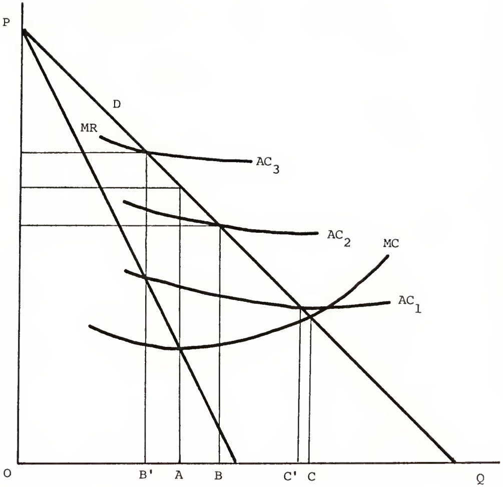


FIGURE 3.1

CONTRACTIONARY OUTPUT EFFECT

constrained firm's output level will not necessarily exceed the monopoly level at output A. Westfield makes the even stronger argument that output may be pushed below A to some level like B' by the rate-of-return constraint. However, he is careful to note that a reduction in output from A to B' would require the unusual condition that capital be an "inferior" factor of production. This situation would cause AC_1 to rise to AC_3, which intersects AR at B'.

A mathematical explanation for the reduction in output below the monopoly level was independently developed by Westfield [p. 430] and Baumol and Klevorick [p. 178]. Figure 3.2 shows the graphical explanation, which has not yet appeared in the literature, as a simple extension of the Zajac framework. Capital is shown to be a regressive factor[8] in Figure 3.2, causing the firm's expansion path to bend backward upon itself. Regulation encourages the firm to reduce its profit-maximizing output level from Q_3 to Q_2. If rays were drawn from the point of origin to the π_{max} and K_{max} positions, it would be clear that the former position lies on the steeper ray. This means that the firm's K/L ratio is smaller at the higher output Q_3. When capital is an inferior factor, then, the simple rate-of-return constraint leads the firm to curtail production while employing more units of capital and fewer units of labor. Although this situation is theoretically

---

[8] An inferior factor has received the name "regressive factor" from Hicks [pp. 93-96], because the negative output effect of the factor indicates a "regression relation."

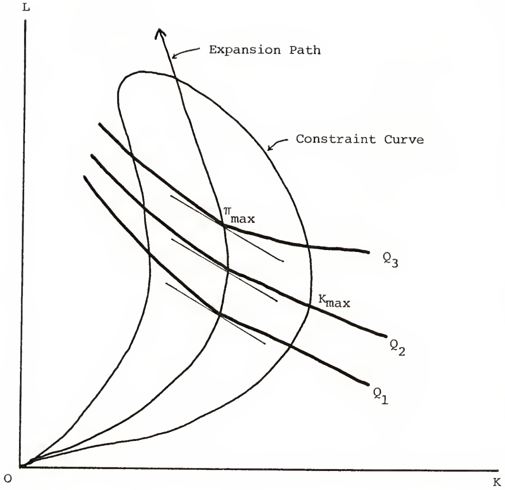


FIGURE 3.2

CONTRACTIONARY OUTPUT EFFECT WITH INFERIOR FACTOR possible, casual empiricism leads Westfield to agree with Baumol and Klevorick that it is most improbable.

The Expansionary Output Effect

The simple rate-of-return constraint may provide the regulated firm with an incentive to sell a portion of its output at below-MC prices. Unless demand is perfectly inelastic, the abnormally low prices will increase the quantity of product that is demanded from the firm and will thus have an expansionary influence on output. Output expansion, in turn, yields the desired growth of the capital rate base. In the absence of the contractionary and expansionary output effects, the regulated firm would produce at the socially optimum level of output. In the presence of the contractionary effect above, output would be suboptimal. The expansionary effect, then, may serve to return output to the desired level or--if the effect is strong enough--push output beyond the socially optimum level. Even if the expansion only serves to return output to the desired level (thus offsetting the contraction exactly), the expansion is undesirable since it can be acquired only through a net loss of consumers' surplus, a result of below-MC pricing. These observations will be explained shortly.

It should be kept in mind that the "expansionary output effect," as it is used here, refers only to the influence on production of below-MC pricing. The effect does not include the influence of price reductions from the monopoly level down to the competitive level (where P = MC). This would be a desirable outcome of regulation. Moreover, the effect does not include the influence of below-AC discriminatory pricing that is needed to allow the firm to break-even, as long as all prices are equal to or greater than MC. In this way the expansionary output effect is defined to encompass only those influences resulting from below-MC pricing. Even within this narrow definition, however, the effect can be subdivided into two influences, one which results from (below-MC) positive pricing and a second which results from (below-MC) negative pricing. An explanation of these two expansionary output effects will now be offered.

Expansion with Positive Marginal Pricing

The first analysis of the expansionary output effect of positive, below-MC pricing by the constrained firm appears in the original Averch and Johnson article [pp. 1057-67].

AJ and below-MC pricing

The economic theory which Averch and Johnson employ in their discussion of the expansionary output distortion is simple and not well-developed. Within the framework of Figure 3.1 above, A and J argue that the constrained firm may not be satisfied to produce output B. Instead, the firm has an incentive to enter marginal markets that would be unprofitable to the unconstrained firm. If the monopolist operates in two regulated markets, for example, and if the rate of return is averaged over both markets taken together, the firm may operate at point A in Figure 3.1. The ensuing monopoly profits could then be used to finance expansion into the second market at a price below marginal cost. The regulated firm can sustain long-run losses in the second market, because its combined profits are at an allowable level. Moreover, the second market provides further expansion of the firm's capital rate base, which justifies a higher profit level.

Averch and Johnson argue that the result of this third-degree price discrimination is the reduction of output in the first market--perhaps a decrease to the monopoly level--and an expansion of output in the second, more price-elastic market. Unfortunately, A and J fail to say that the effect is on the combined output of both markets taken together.[9] For this reason, it is not made clear whether the firm's total output will exceed or fall short of the socially optimal level. Because the firm is now assumed to use third-degree price discrimination, which makes it possible for the firm to equate marginal price and marginal cost in Figure 3.1, the Pareto-optimum level of output shall be here identified as C instead of C'.[10] In the presence of the input distortion, however, the socially desirable (second-best) level is output B instead of C. The important point being made here is this: a return to optimal level C is socially desirable only if the AJ input effect is eliminated, which would return average costs from AC_2 to the efficient level at AC_1. It should be noted, however, that this conclusion is implicit rather than explicit in the AJ analysis.

---

[9] The optimum level of total output is only meaningful, of course, as long as the same product is sold in both markets.

Although A and J do not actually conclude that the firm's use of below-MC pricing causes total output to expand, this conclusion can be deduced from their statement that below-MC pricing in the second market will expand the firm's rate base. If curve D in Figure 3.1 is the firm's total market demand, the firm will sell output B using price discrimination only to recover losses on marginal units sold at marginal cost. However, if the firm finds it profitable to subdivide its total market into two markets, selling at a price below MC in the second market, the combined output must necessarily increase above output B. Since the firm can earn the allowable, competitive rate of return at output B, it would not have found below-MC pricing in a submarket profitable if output (and thus the rate base) were not expanded beyond B. The use of below-MC pricing, then, necessarily implies an expansionary (total) output effect.

---

[10] While output C is the socially optimum level of production, the use of price discrimination violates the exchange efficiency condition that the MRS for any two products must be equal for any consumers consuming both. There is apparently no practical solution to this problem. One remedy would require that the producer's loss at output C be subsidized by a lump sum tax rather than cross-subsidized by price discrimination. However, any feasible tax would create welfare losses of its own. For a discussion of welfare optimization, see Reder [Chapter 2].

Because A and J do not clearly state that below-MC pricing will expand total output beyond level B, it is even less clear whether this expansionary effect will be great enough to return total output to the previously-optimal level at C. Actually, the price-elasticity of demand may be so low in the first market and so high in the second that below-MC pricing in the latter market pushes total output beyond Pareto-optimum level C. The explanation of this result will be presented below with the introduction of the Kafoglis analysis. The point to be made here, however, is that below-MC pricing is socially "detrimental" because it expands output beyond position B. The contractionary output effect of the input distortion is simply offset by the expansionary effect of below-MC pricing. The common wisdom that "two wrongs don't make a right" in this context means that two distortions are not better than one, even when output is returned to Pareto-optimum. Of course, if output is expanded beyond Pareto-optimum, the loss of social welfare is exacerbated. Because this conclusion also is not made clear in the AJ analysis, confusion concerning this result subsequently surfaces in the literature.[11] However, it is best to postpone discussion of this misunderstanding to the end of this chapter. It can be concluded here that, even though Averch and Johnson were the first to identify the below-MC pricing distortion of output by constrained firms, their theoretical analysis of this output distortion is not nearly as strong as their analysis of the input distortion.

---

[11] See Khan [1968 and 1971].

Wellisz and below-MC pricing

Like Averch and Johnson, Wellisz is a pioneer in the identification of regulation-induced input and output distortions. The Wellisz paper was developed independently of the AJ article, and the former is preceded in print by the latter by only two months. Where the AJ theoretical framework scrutinizes the input distortion but only weakly explores the output distortion, the Wellisz analysis--by contrast--is weak with respect to the input distortion but strong with respect to the output distortion. In fact, the Wellisz discussion of input distortions is so weak that he gives it little attention--and only then when it is discussed as incidental to the output distortion.[12]

A and J had shown that the constrained firm will use monopolistic pricing in the first market as long as price in the second market does not fall below MC. On the other hand, Wellisz uses a peak-load pricing model which is free of any price discrimination. Within this framework, Wellisz shows that a simple rate-of-return constraint will induce the firm to always use monopolistic pricing in its off-peak market. The monopoly profits on the off-peak sales can be used to finance the deficit on the peak-load sales. In this way the firm can drive its peak-load investment beyond the level at which the allowable return from peak-load sales would be earned.

---

[12] Because the literature largely overlooks the importance of the expansionary output distortion, and because Wellisz is mainly concerned with this output effect, his analysis is generally relegated to footnote status in articles devoted to the AJ input effect. An exception is Kahn's use of the term, "A-J-W effect" (so named to include Wellisz). Unlike most economists, however, Kahn associates A and J quite equally with both input and output distortions (although it is not always clear which association he intends at the time he uses "A-J-W effect"). See Kahn [1968 and 1971].

If average costs are constant or rising, this cross-subsidization enables the peak-load customers to purchase the firm's service (or product) at a price below opportunity costs, and output is overexpanded. If the average cost curve declines, Wellisz [p. 41] argues, the price paid by peak-load customers may be greater than, equal to, or less than opportunity cost, depending upon the degree of scale economies. Regardless of the shape of the cost curve, however, Wellisz shows that the constrained firm will increase profits by setting off-peak prices at the monopolistic level and using some of the earnings to help finance peak-load costs. As a result, Wellisz is able to conclude that "while the imposition of a 'fair-return' limitation favors the peak-load customer, it does not protect the off-peak, purchasers of service" [p. 37].

Wellisz's use of "below-opportunity cost" pricing easily translates into the below-MC pricing practice identified by Averch and Johnson. This transition in terms can be made because the opportunity cost of servicing off-peak customers is short-run marginal cost, and the opportunity cost of servicing peak-load customers is long-run marginal cost (which includes the cost of capacity demands). Although the deleterious effect of below-MC pricing is identified by A and J within a third-degree price discrimination model and by Wellisz within a peak-load pricing model, the strongest case against below-MC pricing is later made by Kafoglis within a first-degree price discrimination model.

Kafoglis and below-MC pricing

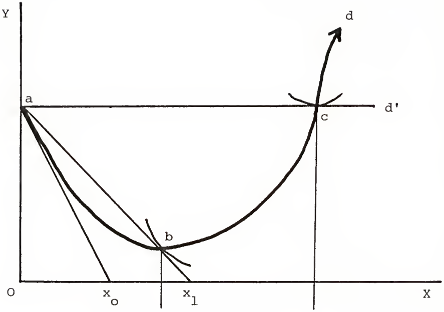
It may be remembered from Chapter 2 that Westfield carried the AJ input analysis to its logical extreme by dropping the AJ assumption that the MPP_K be positive. In a similar way, Kafoglis carries the AJ (expansionary) output analysis to an extreme by dropping the restriction that price discrimination is only of the third degree. Kafoglis employs a first-degree price discrimination model, which is capable of defining the outer limit of output distortion that results from below-MC pricing. Actually, Kafoglis extends the AJ output analysis only to its "probable" extreme--not to its logical extreme--as will be shown later.

While important similarities exist between the AJ and Kafoglis studies of output distortion, an important difference also can be found. While Averch and Johnson's constrained firm is a profit-maximizer, Kafoglis' firm is an output-maximizer, whose objective is to maximize output while earning the fair rate of return. This difference in objectives has its most striking effect on the two models when the fair rate of return is set equal to the market cost of capital. This means that the firm's profits are zero, so that the AJ model--unlike the Kafoglis model--would be unable to yield a determinate solution[13] (unless, of course, a secondary objective were added). However, when the rate-of-return constraint exceeds the cost of capital but is low enough to be restrictive, the Kafoglis model can easily be used to extend the AJ output analysis. The output-maximizing objective need only be replaced by profit-maximization. In this way, the ill effects of below-MC pricing on output can be explored within the theoretical framework developed by Kafoglis.

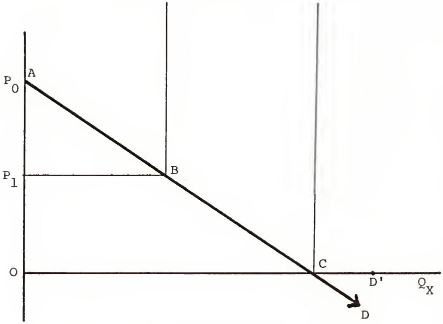
Average revenue without price discrimination.--Before applying the Kafoglis analysis to the behavior of A and J's constrained profit-maximizer, it will be helpful to briefly examine the firm's average revenue curve while the complications of price discrimination are still absent. A linear demand curve, line ABCD, is derived for a single-market firm in Figure 3.3. Goods X and Y are substitutes for each other.

---

[13] Averch and Johnson [p. 1055] recognized that their model would be indeterminate in this situation. For a more recent discussion of the problem, see Baumol and Klevorick [p. 174].



(a) Price Consumption Curve



(b) Demand Curve

FIGURE 3.3

DERIVATION OF THE FIRM'S DEMAND CURVE

The price is determined exogenously for unregulated product Y but endogenously for regulated good X. Budget line ax_0 shows the alternative combinations of X and Y that the consumer community can buy when the price of X is P_0 in Figure 3.3(b). At this price the quantity demanded of X is zero, which means that the budget line is tangent to the community indifference curve (not shown) at point a in Figure 3.3(a). Budget line ax_1 shows the alternative combinations of X and Y when the price of X is P_1, and budget line acd' corresponds to a zero price of X.

The product demand curve is derived from the community's price consumption curve abcd'. The elastic portion AB of the demand curve, for example, is derived from the negatively sloped section ab of the PCC. Similarly, the inelastic segment BC of the demand curve is obtained from the positively sloped portion bc of the PCC. Finally, the community indifference curve is assumed to be tangent to budget line acd' throughout the infinitely long segment cd' of the PCC. The infinitely long segment CD' of the demand curve is derived, in turn, from segment cd' of the PCC. If the regulated firm does not use price discrimination (and is not subsidized), it will operate at that level of output where its downward sloping AC curve intersects the AR curve.

Average revenue with price discrimination.--There is no reason, however, why price discrimination will not be employed, since it is not proscribed by simple rate-of-return regulation and since it is clearly profitable to the constrained firm. This is the main point made by Averch and Johnson in their discussion of output distortions by the multimarket utility. Kafoglis is the first to carry this reasoning to its "probable" extreme and the first to present the argument within a single-market framework. His approach can be explained using Figure 3.4, though the Kafoglis model actually employs total instead of average cost and revenue curves.[14]

Without price discrimination, the constrained firm would produce q_0 units of X, the level at which AC equals demand. At this point, D would still represent the AR curve. With the use of first-degree price discrimination, however, the firm's AR curve swings outward from curve Pq_3 to a new position at Pq_5. The demand curve, remaining stationary, is now the MR curve. The firm breaks even, earning its allowed, competitive rate-of-return at output q_2. Production units 0 to q_0 are sold at prices above AC and MC and production units q_0 to q_2 are sold at prices below AC and MC. The expansionary output effect in this less successful case would be distance q_0q_1.

Expansion with Negative Marginal Pricing

It was shown in Chapter 2 that Westfield carries the AJ input analysis to its logical extreme. It remains to be shown here, however, that Kafoglis (using a different

---

[14] Although the average curves do not appear in his article, their use was suggested to me during a class lecture by Kafoglis.

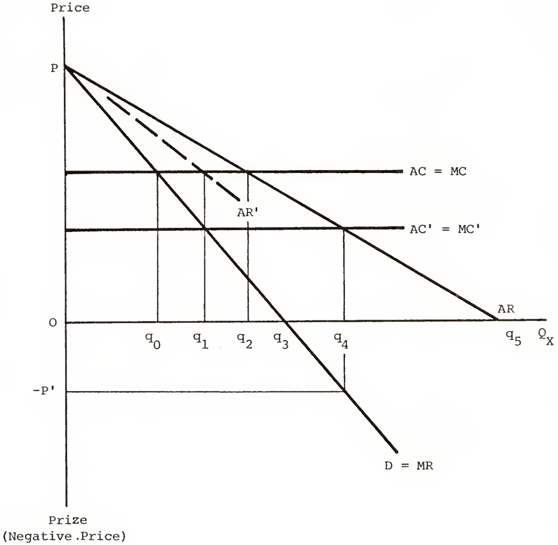


FIGURE 3.4

BREAK-EVEN SOLUTION WITH PRICE DISCRIMINATION maximand) carries the AJ, below-MC pricing argument only to its "probable" extreme. Had Kafoglis actually carried the output analysis to its logical extreme, he would have considered the possibility that the constrained, price-discriminating firm not only employs below-MC pricing--but below-zero pricing as well. When carried to this (perhaps, improbable) extreme, the argument that simple rate-of-return regulation can induce an expansionary output distortion is strengthened considerably.

It is reasonable that Kafoglis would ignore the negative marginal pricing possibility if he were implicitly assuming that the demand curve in Figure 3.4 intersects points P, q_3, q_4, and q_5. While it is true that this assumption would effectively eliminate the possibility of negative pricing, it would require that the TR curve be horizontal for all outputs above q_3. However, the Kafoglis model [p. 588] shows the TR curve (which is produced by first-degree price discrimination) to be rising, not horizontal. Note that q_3 lies below the midpoint of the discriminating AR curve. Since Kafoglis examines only the rising portion of the perfectly discriminating TR curve, he considers only the upper half of the AR curve shown in Figure 3.1.[15] In this way, the entire lower half of the AR curve is overlooked. Since it is possible that average cost is as low as AC' and that demand is represented by the straight line $Pq_3q_4q_5$, it could be profitable for the restrained firm to give production

---

[15] As was noted earlier, Kafoglis' "Output of the Restrained Firm" extends the AJ output analysis (using a different maximand) in much the same way that Westfield's "Regulation and Conspiracy" extends the AJ input analysis. There are other similarities between these two articles, which might appear to be concerned with disparate concepts, that are quite striking. For example, unlike the average cost

---

and revenue functions used here in Figure 3.4, total cost and revenue functions are used in both the Kafoglis model and Westfield model. Kafoglis' Figure 2 [p. 588] is to output analysis what Westfield's Figure 2 [p. 434] is to input analysis.

Unlike Westfield's Figure 2, however, Kafoglis' Figure 2 contains a technical error that prevents the author from carrying the AJ output analysis to its logical extreme. This error can be quickly identified with reference to Figure 3.4 above. Kafoglis' Figure 2 uses total cost and revenue functions, but the average cost and revenue functions in Figure 3.4 are drawn consistent with his assumptions of constant cost and a linear demand curve.

Kafoglis' analysis uses three TR curves. The first TR curve, which assumes no price discrimination, and the second TR curve, which assumes third-degree discrimination, are shown to be falling when intersecting the TC curve. In terms of Figure 3.4, this means that average cost must be curve AC' which intersects D and AR' below their midpoints. It must follow, then, that AC' also intersects curve AR (which assumes the presence of first-degree price discrimination) below its midpoint. This means that MR must be negative regardless of the degree of price discrimination employed by the firm. Accordingly, Kafoglis's third TR curve, drawn under the assumption of first-degree price discrimination, whould be descending like the other two TR curves when it intersects the TC curve. Unfortunately, the third curve is shown to be rising at the point of intersection. Had this curve been shown to descend, as it must, it would be obvious that MR is negative and (because D = MR here) that marginal prices are negative also.

In contrast, Westfield builds his entire conspiracy argument upon his recognition that the Total Net Revenue Product of Capital Curve (T_K) might be falling when it intersects the Total Capital Cost Curve (C_K). This use of T_K and C_K in factor distortion analysis is analogous to Kafoglis' use of TR and TC in his discussion of output distortions. Westfield [pp. 438-40] even acknowledges that market saturation could account for the decline in T_K as well as physical saturation, but he believes the former to be much less likely than the latter. Even though Westfield does recognize the efficacy of negative MR, he does not envision negative marginal pricing, since he does not explore price discrimination and output distortions.

units q_3 through q_4 away free of charge.

The Kafoglis analysis can be carried one important step further if the demand curve is assumed to continue downward from the first quadrant in Figure 3.4 into the fourth quadrant. In the earlier Figure 3.3, this would mean that demand is represented by the straight line ABCD, which is linear throughout its relevant range. The price consumption curve would now become curve abcd instead of curve abcd'. In Figure 3.3(b), it can be seen that the firm must give its marginal customers prizes (i.e., charge negative marginal prices) to force the quantity of product demanded to exceed point C.[16]

Returning to Figure 3.4, it is evident that production units zero through q_1 are sold at above-AC prices. Units q_1 through q_3 are dumped onto the market at below-MC prices. The q_3th unit is given away free of charge. And, finally, units q_3 through q_4 are forced into a saturated market because the reluctant customers are bribed with prizes worth as much as P'. The negative marginal pricing will occur anytime the AC curve intersects the discriminating AR curve below its midpoint. As discrimination efforts become less successful, the AR curve swings back toward its position of coincidence with D. As this happens, the critical lower portion of the

---

[16] Consequently, the demand curve is "prize inelastic" throughout its infinite length below the X axis. The prize elasticity coefficient, which is zero at point C, asymptotically approaches unity as the value of the prize increases to infinity.

AR curve (which makes negative marginal pricing possible) becomes an increasingly smaller proportion of the curve. When this critical portion is very small, the chance that it will be intersected by the AC curve is greatly increased if the AC curve is downward sloping, as would likely be the case for a regulated utility. Negative marginal pricing becomes more likely, then, as second-degree price discrimination approaches first-degree and as the AC curve descends more rapidly from the vertical axis.

Actually, the constrained firm would find it more profitable to increase average costs (by substituting capital for labor) than use negative marginal pricing. If average costs are at AC' in Figure 3.4, for example, allowable profits would be maximized when the firm raises curve AC' to the level of curve AC. This result is deduced from the Westfield analysis [p. 439], which argues that market saturation will cause the firm's T_K curve[17] to peak and descend. With the occurrence of market saturation, the firm would substitute capital for labor, driving up costs until output is reduced to the highest level at which market saturation does not occur. Should market and physical saturation exist simultaneously, the firm would find it profitable to purchase nonproductive capital (instead of substituting productive capital for labor). The acquisition of useless capital[18] would raise average costs, causing output to fall until output markets just cease to be saturated. Westfield is thus able to conclude that "the firm will dodge market saturation so that physical saturation will occur first" [Westfield, p. 429].

---

[17] Westfield's terminology, including the "T_K curve," are explained earlier in Chapter 2.

An important point in Westfield's argument is that simple rate-of-return regulation cannot alone force the firm to operate on the inelastic portion of its demand curve. If the firm finds its average cost curve to intersect the inelastic region of its demand curve, it will simply raise the average cost curve (by acquiring capital) until it intersects the demand curve at the point of unit price elasticity.[19] This means that the firm would be found operating on the inelastic portion of its demand curve only if (1) market saturation occurs and (2) regulators can effectively police the firm's operations to prevent it from inefficiently raising its average costs (by means of capital acquisition). Negative marginal pricing, then, will occur only when these two unlikely circumstances are both found to exist. Even though the firm's average total cost curve is not likely to intersect the inelastic portion of its demand curve in reality, this improbable situation is commonly assumed in the theoretical models of economic literature. Within the field of regulatory economics alone, the average total cost curve is shown to intersect the inelastic portion of the demand curve in models used by Averch and Johnson [p. 1057, Figure 2], Bailey and White [pp. 9,15,18 and 27, Figures 1-4], Kafoglis[20] [p. 584, Figure 1], and McNicol[21] [p. 435, Figure 3]. Within the context of these theoretical models, at least, negative marginal pricing is a viable policy of the constrained firm. This possibility should no longer go unrecognized.

---

[18] Alternatively, the firm could conspire with its capital supplier to raise the cost of productive capital. This is effectively the same thing as purchasing useless capital.

[19] Although Westfield's proof is informal, Bailey [p. 107] presents a mathetical proof that the firm constrained only by rate-of-return regulation will not operate on the inelastic portion of its demand curve.

Negative marginal pricing has thus far been described as a logical--though improbable--extension of the AJ analysis of output distortion. Even though it may appear to be an improbable situation, one can still wonder what the price effects are of promotional prizes like free internal wiring for gold medalion homes and free gas lamps for natural gas consumers. Whatever its empirical significance is found to be, the fact that a regulated firm could increase profits by dumping a portion of its output at below-zero prices certainly increases the plausibility that below-MC prices are a reality. The Averch-Johnson, Wellisz, and Kafoglis discussions concerning below-MC pricing may have earlier appeared alarmist and far afield. When compared with the strange results of the present analysis, however, their more conservative warnings appear more realistic and closer to home base. Simple rate-of-return regulation can lead to serious, expansionary output distortions, indeed.

---

[20] Note that AC intersects D to the right of the intersection of MR with the horizontal axis.

[21] McNicol's Figure 3 [p. 435] can be included in this group if he intends, as seems likely, that cost curve C' represents both marginal and average costs. This interpretation is suggested by the graph's title, which includes the term, "average cost." Furthermore, no fixed costs are specified for the model, which suggests that average costs are intended to equal the constant marginal costs.

Output Effects and Their Analogs

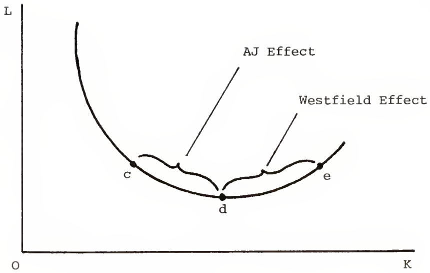
A symmetry can be observed between the output distortions identified in this chapter and the input distortions reviewed in Chapter 2. This should not be surprising, since the use of isoquants in production theory is simply an application of the use of consumer indifference curves. Three analogous concepts will now be identified through the comparison of (1) isoquants with indifference curves, (2) the expansion path with the income consumption curve, and (3) the cost employment curve with the price consumption curve. This analysis will demonstrate that the AJ input effect is an analog of the KW output effect and that the Westfield input distortion is an analog of negative marginal pricing (here to be named the "quasi-Westfield effect").

Isoquants and Indifference Curves

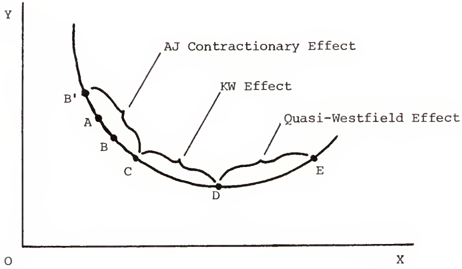
Because of the input and output distortions, the regulated firm purchases too much capital (relative to labor) and the consumer purchases too much of product X (relative to unregulated product Y). The firm's purchases of capital are distorted, because the effective cost of capital is reduced by capital's "premium," the nonproductive value of capital as a rate base and reducer of excess profits. Similarly, the consumer's purchases are distorted because the marginal price of the product is reduced by cross-subsidization. Because "capital premium" is an unfamiliar term, it will be explained below. Following this explanation, the input-distorting influence of the capital premium will be demonstrated with isoquants, and the output-distorting effect of cross-subsidization will be shown with indifference curves. In this way, the distorting effects of the capital premium and cross-subsidization will be shown to be analogous.

Isoquants and the capital premium

The firm's effective cost of capital is less than the market cost by the amount of the nonproductive premium associated with capital purchases. The input distortion caused by this premium is demonstrated in Figure 3.5(a). Points c, d and e on the isoquant in Figure 3.5(a) correspond exactly to points X, A and B which were earlier identified on an isoquant in Figure 2.7. This means that point c must lie on



(a) Input Distortions by a Firm



(b) Output Distortions by a Buyer

FIGURE 3.5

COMPARISON OF OUTPUT AND INPUT DISTORTIONS the expansion path (i.e., the locus of efficient points), if the constrained firm is operating somewhere on the isoquant in Figure 3.5(a). If the point of operation lies anywhere between points c and d, the substitution effect between the point of operation and point c measures the AJ input effect. If the operating position lies anywhere beyond point d (say, at point e), the substitution effect cd measures the AJ effect and the substitution effect de (in this example) measures the Westfield effect.

It may be recalled that the firm will operate at position c in Figure 3.5(a) if its market and effective costs of capital are the same. These costs do differ, however, because K has a value as a basis upon which profits can be justified.[22] If the allowed rate of return on K is 10 percent, for example, the value of capital stock as a rate base is 10 percent of its market cost. In Chapter 2 it was assumed that K and L are both acquired in a competitive factor market. If it were also true that the regulated firm is selling its product X in a competitive market, the firm would purchase additional units of K until the price of K equals the value of marginal product of K (VMP_K) plus the marginal value of the K rate base ($MVRB_K$). That is, the firm acquires K up to the level at which

---

[22] Averch and Johnson [p. 1053] were the first to show that the constrained firm's "private" cost of capital is not equal to the market or "social" cost of capital.

$$P_K = VMP_K + MVRB_K. \qquad (1)$$

Capital can serve the firm in three ways. First, it is productive, and this function of K is reflected in a positive MPP_K, and hence, a positive VMP_K (when P_K is positive). Second, the K rate base is the basis upon which profits are justified. Third, the K rate base is a means by which excess earnings can be reduced. For example, if it is assumed that the MMP_K price is positive, that the price of K is \$5, and that the firm's rate of return is excessive, the $MVRB_K$ can be as high as \$4.50. This follows because the marginal unit of K increases allowable profits by \$0.50 and increases total costs (i.e., reduces excess earnings) by \$5.00. The $MVRB_K$ has two components, the value of K in supporting profits and the value of K in reducing excess earnings. If product demand is so great that the firm must employ capital at a level of negative MRP_K, the $MVRB_K$ can even exceed \$5.50 (i.e., exceed 110 percent of P_K in the above example) since the negative MRP_K is valuable if the firm is to comply with its constraint.

This relationship is more easily seen with additional examples. If the firm is operating at socially efficient point c, the $MVRB_K$ is zero because the firm is unconstrained by regulation. At point c, the \$5 price of K equals the VMP_K. If the regulated firm operates at a position between points c and d, the VMP_K must be some value less than \$5--say, \$3. From equation (1) it is clear that the $MVRB_K$ must be \$2, because total profits would exceed the allowable level by \$2 if the last unit of K were acquired free of charge. Although the market price of K is \$5, the firm's effective price of K is only \$3 (i.e., the \$5 market price less the \$2 MVRB_K). The firm equates the VMP_K with capital's effective price (P_K^*)--not capital's market price (P_K)--to determine the level of capital stock.

If instead the firm were operating at position d, the MPP_K and VMP_K would both equal zero. The price of K, in this case, is fully offset by the MVRB_K so that the firm's effective price of K is zero. When the firm's K_{max} position is at point d the MVRB_K must be \$5. On the other hand, when K_{max} occurs at point e, the MVRB_K must exceed P_K (since the VMP_K is negative) and P_K^* must be negative.

The above examples assume that X is sold competitively, which will not likely be the case for a regulated firm. Since the demand curve for X will be downward sloping, the VMP_K in equation (1) must be replaced by the MRP_K, yielding

$$P_K = \text{MRP}_K + \text{MVRB}_K. \tag{2}$$

Because MR is always less than AR, the MRP_K is always less than the VMP_K. As a result, the MRP_K must be negative when the firm is operating at point d, since the VMP_K would be zero at that position. The MVRB_K is greater than P_K at point d. In fact, the MVRB_K is so great that it exactly offsets both the cost of the last unit of K and its detrimental effect on MR. As a result, the effective price of capital can still be said to be zero at position d, even though MRP_K is now assumed to be negative.

The purpose of this discussion has been to identify the $MVRB_K$ as the premium of marginal units of K that induces the firm to purchase K beyond the level at which $P_K = MRP_K$. This premium has never been clearly identified in the literature. Averch and Johnson have left the impression that capital's only non-productive value is its value as a basis for profits. As has been shown here, the capital rate base is valuable for two non-productive reasons: (1) it supports profits and (2) it reduces unsupported profits. This latter reason has received little attention in the literature, perhaps because its effect is relatively small when the firm's K_{max} position is located near the socially efficient point c. However, when K_{max} moves rightward along the isoquant, capital's value as a profit reducer increases continually.

This reasoning is perhaps most apparent when K_{max} is located at point d. If X is sold competitively, the firm will be enticed to operate at point d only if the premium of the marginal unit of K (i.e., the $MVRB_K$) equals P_K so that P_K^* equals zero. If X is not sold competitively--a more realistic assumption--the firm will not operate at position d unless the premium exceeds P_K by an amount sufficient to offset the negative MRP_K. Whether X is sold competitively or not, the premium of the incremental unit of K must exceed P_K if the firm is found operating at a point beyond d (say, at e).

It may seem strange to conclude that capital can be valuable as a profit-reducer. Nevertheless, this conclusion addresses the question "Why would a regulated firm pay $5 for a unit of capital that is worth only $3 as a factor of production and only $.50 as a profit base?" The answer must be that the unit's $1.50 value as a profit-reducer enables the firm to avoid violation of its constraint. Now that the capital premium has been identified, a direct comparison can be made between its distorting influence and that of consumer cross-subsidization.

Indifference curves and cross-subsidization

In the same way that the capital premium lowers the firm's effective price of K, cross-subsidization (among consumers or consumer classes) lowers the favored consumer's market price of X. This leads both firm and consumer to purchase wastefully large quantities of K and X, respectively. However, an important difference between the two distortions should not be overlooked. While the regulatory constraint distorts the firm's factor preference, leaving the market price of K unchanged, the constraint distorts the market price of X, leaving the consumer's product preference undisturbed.[23] Of course, this distinction does not mean that input and output distortions are not analogous. It only means that they are not identical.

---

[23] The firm's preference function for K and L can change to favor K more strongly even though the isoquant remains unchanged. The isoquant, after all, is determined by a production function--not a preference function.

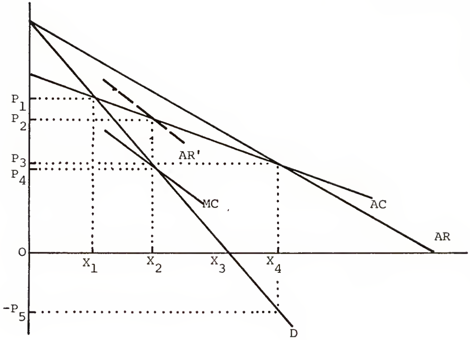
Personal indifference curve.--The analogous relationship between input and output distortions is demonstrated in Figure 3.5. Corresponding to points c, d and e on the isoquant in Figure 3.5(a) are points C, D and E on an individual's indifference curve in Figure 3.5(b). This individual is assumed to be the regulated firm's marginal customer, so that he is always quoted the firm's lowest price. The socially efficient quantity of X is demanded by the entire community when $P_X = AR_X$ and the marginal price of $X = MC_X$. It is also assumed that the X producer enjoys scale economies and that society prefers to subsidize its operations using consumer class cross-subsidization rather than government subsidization. The consumer in question is a competitive buyer, so the firm's lowest price applies to all of his purchases, which are a negligible portion of the community's purchases of X. Finally, it must be assumed that good Y is sold competitively, so that its price is equal to its AC and MC.

In the absence of the AJ contractionary output effect, the regulated firm might equate the marginal price of X with MC_X so that the marginal buyer selects position C in Figure 3.5(b). Point C, then, is the socially efficient point of consumption and, as such, is analogous to point c in Figure 3.5(a). If the marginal P_X is less than MC_X, the resulting subsidy will lead the buyer to favor a lower marginal rate of substitution of X for Y, moving him away from point C. The buyer will be found in consumer equilibrium at point D only if the marginal $P_X = 0$ and at position E only if the marginal $P_X < 0$. Arc CD, then, can be identified as the KW effect and arc DE as the quasi-Westfield effect.

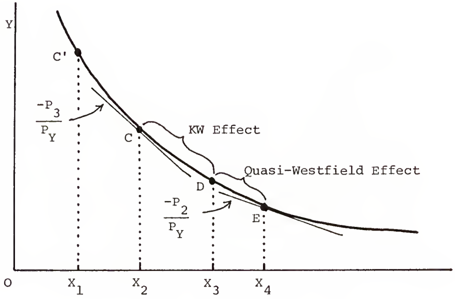
In the absence of the KW effect, the AJ contractionary output effect could lead the firm to charge such a high marginal price that the consumer's location might be found at points B, A or B' (as explained earlier with Figure 3.1). If the contractionary effect is present, then, the expansionary KW effect must be represented by a substitution effect greater than CD if the buyer is found to maximize satisfaction at point D. For example, if the contractionary effect were arc BC and the consumer were located at point E, the KW effect would encompass the segment BCD and the quasi-Westfield effect would include segment DE. Whether the AJ contractionary output effect is present or not, the AJ and KW effects are shown to be analogous, and the Westfield effect is revealed to be the input counterpart to negative marginal pricing (the "quasi-Westfield effect").

Community indifference curve.--It has been shown in Figure 3.5(a) that the firm can operate on the horizontal portion (point d) of its isoquant because the capital premium can fully offset the cost of capital (including its price and detrimental effect on TR). Similarly, it has been shown in Figure 3.5(b) that the individual can locate on the horizontal portion (point D) of his indifference curve, because the cross-subsidy from other consumers may fully pay his bill. For the product market as a whole, however, no subsidization exists. Even though the regulated firm is willing to charge favored customers below-MC prices--or negative prices, even--it is unwilling to supply the entire community of consumers with a product having an average price below average cost. For this reason, the relationship between the community indifference curve and output effects is dissimilar to that between the personal indifference curve and output effects. Consequently, a few remarks concerning the community indifference curve are now warranted to supplement the foregoing discussion of the personal indifference curve.

Figure 3.6 illustrates the effect of output distortions on the equilibrium position of buyers located on the community indifference curve. In part (a) of the figure, buyers will all be charged a high price of P_1 if the firm just covers AC without use of price discrimination. In this case, quantity demanded is X_1 and the buyers will locate at position C' on the community indifference curve in part (b) of the figure. The community would be better served, however, if the firm would use second- or third-degree price discrimination, charging prices as low as P_4--which equals MC--to marginal customers. This judicious use of discrimination would cause the AR curve to separate from the D curve and swing outward



(a) Community Revenue Curves



(b) Community Indifference Curve

FIGURE 3.6

COMMUNITY OUTPUT DISTORTIONS

to the dashed position at AR'. Accordingly, the buyers' average price would now fall from P_1 to P_2, causing an increase in the consumption of X to the socially efficient level X_2. Community consumer satisfaction is maximized at point C.

Even so, the constrained firm might rather hope to maximize profits by using first degree price discrimination. In this situation, the AR' curve swings all the way out to the position identified as the AR curve, and average price falls to P_3. Marginal price is now $-P_5$, quantity demanded is X_4, and consumers' equilibrium is now located at point E. If, instead, AC had intersected AR at the midpoint of the latter, the marginal price would have been zero and consumers' equilibrium would have been situated at point D. Consequently, this analysis identifies segment CD as the KW effect and segment DE as the quasi-Westfield effect.

If the situation were complicated by the addition of the AJ contractionary output effect, the cost curves would rise and points C and E would slide upward along the indifference curve. Of course, point D would remain unchanged, so the potential KW effect would increase and the quasi-Westfield effect would diminish. If costs were increased sufficiently, the latter (and even the former) effect could be eliminated entirely.[24]

---

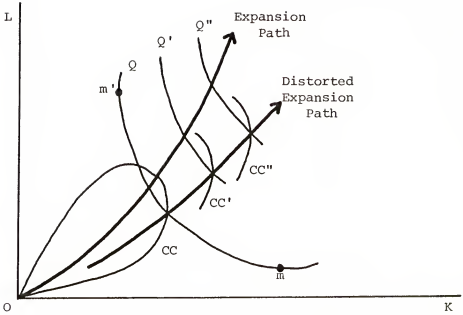
[24] In this way, an increase in the input distortion would reduce the combined output distortion. This interesting trade-off will be more fully explained in Chapter 4.

Expansion Path and Income Consumption Curve

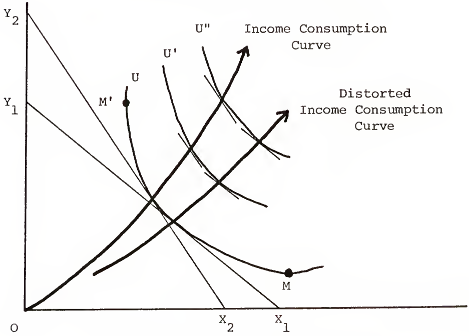
The analogy between output and input distortions can be strengthened by a comparison of the firm's expansion path with the consumer's income consumption curve. Figure 3.7(a) illustrates the firm's expansion path, a locus of efficient points drawn under the assumption that input prices are given and that product demand can be at any of many levels. Lying to the right of the expansion path is the "distorted expansion path,"[25] which is a locus of all the K_{max} points that would be produced by variations in product demand. As an example, increases in demand could expand the constraint curve CC outward to position CC' or CC". The three alternative K_{max} positions, in this case, are the points of intersection of isoquant Q with constraint curve CC, Q' with CC', and Q" with CC". Although they are not shown, ridge lines would intersect capital's intensive margin for output Q at point m and capital's extensive margin for output Q at point m'. Production stage I lies to the left of the ridge line through point m', and production stage III lies to the right of the ridge line intersecting point m. The middle ground, of course, is the "economic" stage II. As product demand grows, the unconstrained monopoly follows the expansion path outward and the regulated firm follows the distorted expansion

---

[25] This distorted expansion path is identified by Averch and Johnson [p. 1053] as "expansion path 2." In the field of international trade, Magee [p. 18] identifies a similar path which he calls the "distorted efficiency locus."



(a) Efficiency Locus for Firm



(b) Efficiency Locus for Consumer

FIGURE 3.7

COMPARISON OF FIRM AND CONSUMER EFFICIENCY LOCI path. At any given level of output, the input distortion can be measured as the length and slope of a line connecting the two points of intersection of the isoquant with the two expansion paths.

Growth in consumer income could easily account for the increases in product demand and (in turn) factor demand that are needed to produce the firm's two expansion paths (normal and distorted). Likewise, an increase in consumer income could produce the two income consumption curves in Figure 3.7(b). Because good Y is sold competitively, the socially efficient income consumption curve (ICC) will be effective only when the average price of X produces a budget line parallel to budget line Y_2X_2. The proportions in which X and Y are consumed will be distorted by the regulated X firm's use of below-MC pricing. Consequently, a distorted ICC is created that relates to the normal ICC in a manner analogous to the way in which the distorted expansion path relates to the unconstrained expansion path. The introduction of simple rate-of-return regulation in this example fosters below-MC pricing and induces consumers to relocate from positions on the ICC to positions on the distorted ICC.

If indifference curve U in Figure 3.7(b) reflects an individual's preference function, the distorted ICC may be found in "consumption stage III," which is not unlike finding the distorted expansion path in production stage III. For a given level of unility U, ridge lines intersect points M and

M', defining good X's intensive and extensive consumption margins, respectively. Consumption stage I is the region to the left of X's extensive margin, stage II lies between the ridge lines, and stage III occupies the area to the right of X's extensive margin. On the other hand, when community indifference curves are under consideration it is clear that the distorted ICC will never enter consumption stage III. The community will not be confronted with even below-AC prices, much less prizes or other inducements. For the favored individual, however, stage III of consumption is a real possibility that has never been acknowledged in the regulatory or general economic literature. This neglect can probably be linked to a similar omission of negative marginal pricing. Corresponding to the measure of input effects, the measure of output distortions is the slope and length of a straight line connecting the two points of intersection of the indifference (iso-utility) curve with the normal anddistorted income consumption curves.[26]

---

[26] The normal and distorted efficiency loci in parts (a) and (b) of Figure 3.7 will intersect one another at some point. In part (a), if demand is reduced sufficiently the area bounded by the constraint curve will eventually be reduced to a single profit-maximizing point on the firm's expansion path. The expansion paths of the constrained and unconstrained firms are identical up to that point, indicating that regulation is effective (and input distortions are produced) only at higher levels of output and demand. In part (b), the normal and distorted income consumption curves will also coincide at low levels of utility if price discrimination cannot be used when consumers have small budgets. On the other hand, if price discrimination is always used, the distorted income consumption curve will intersect the normal curve at the point of origin.

Cost Employment and Price Consumption Curves

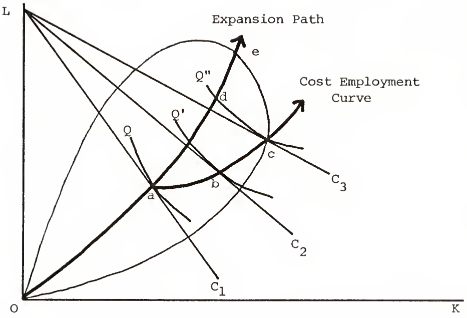
Input and output distortions are analogous in yet another way, which shall now be demonstrated using Figure 3.8. In part (a) of this figure, the firm is confronted with isocost curve C_1 if regulation is so lax that the firm's effective and market costs of capital are the same. The (effectively) unconstrained firm will operate at the profit-maximizing position a on the expansion path. No input distortion yet exists. However, as regulation becomes more and more restrictive the maximum allowable rate of return on capital diminishes continually. This causes the constraint curve, which initially is only the size of point a, to expand outward so that the moving K_{max} position traces out curve abc. This locus of K_{max} points,[27] which heretofore has gone unrecognized, might best be called the "cost employment curve," since it reveals the firm's choice of factor employment at various effective costs of K.

Cost employment curve

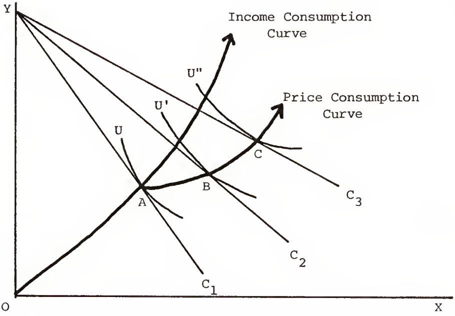
At points a, b and c on the cost employment curve isoquants Q, Q' and Q" are tangent to isocosts C_1, C_2 and C_3, respectively. The cost employment curve, then, shows that

---

[27] Note the difference between the distorted expansion path and the cost employment curve: while each is a locus of K_{max} points, the former is produced by variations in final demand and the latter by variations in effective cost of capital. Furthermore, the distorted expansion path is terminated when K and L are both exhausted, while the cost consumption curve terminates when it reaches the base of the profit hill.



(a) Cost Employment Curve



(b) Price Consumption Curve

FIGURE 3.8

PRICE CONSUMPTION CURVE AND ITS ANALOG

stronger regulation with lower allowable rates of return induces the firm to depart further from the efficient expansion path, increasing the input distortion. Ideally, regulation would force the firm to expand operations from point a to socially efficient point e. However, the AJ contractionary output effect serves to reduce the firm's efficient level of output at position e to output Q" at point c. Because Q" would be produced most efficiently at point d, it is implied by Zajac [1970, p. 120] that arc dc is the AJ input effect at this level of output. It is apparent from Figure 3.8(a) that the graphical representation of the AJ input effect is simply a clever application of Hicks' factor substitution effect.[28] As regulation reduces the effective cost of capital, the output effect moves the firm from point a to point d and the substitution effect relocates the firm's point of operation to position c.

The cost employment curve is a simple geometrical device that can be used to prove that the closer the allowed rate of return gets to the market cost of capital the greater is the input distortion and the larger is the capital stock.[29]

---

[28] See Hicks [Chapters 1 and 2].

[29] The cost employment curve also quickly identifies capital as a normal or inferior factor. If capital is inferior, the cost employment curve lies below isoquant Q in Figure 3.8(a). This is consistent with the earlier analysis of an inferior factor in Figure 3.2. Moreover, Figures 3.8(a) and 3.2 are the first graphs used to explain the effects of an inferior factor on a regulated firm. These results were first identified verbally and mathematically by Westfield [p. 430] and later by Baumol and Klevorick [p. 178].

This result is explained by the firm's outward movement along the cost employment curve as the constraint becomes increasingly more restrictive. This conclusion is more fully explained by Baumol and Klevorick [pp. 174-75] and Scherer [p. 533]. In addition, a rigorous mathematical proof is offered by Takayama [pp. 255-60]. Interestingly enough, the only previous geometric proof is Westfield's rather unorthodox model [pp. 429-32]. The cost employment curve, which explains the result expeditiously within the context of the familiar Zajac model, has apparently been overlooked.

Price consumption curve

The analog of the cost employment curve in part (a) of Figure 3.8 is the price consumption curve in part (b). The consumer (whether it be an individual or the community) is assumed to maximize satisfaction at point A when the marginal price of X equals MC and the socially desirable quantities of X and Y are demanded. Money income and P_y are fixed. As the allowable rate of return on capital is lowered, the firm may find it necessary to lower the marginal price of X below MC to eliminate excessive earnings. Below-MC pricing could have the effect of swinging the isocost curve outward to C_2 or C_3. In this way, the price consumption curve is traced through points A, B and C. If the buyer moves from point A to C, the move from A to D is the income effect and the move from D to C is the substitution (of X for Y) effect. If the isoutility curve U" belongs to an individual, substitution effect DC in Figure 3.8(b) measures the KW effect. If the isoutility curve U" is instead a community indifference curve, arc DC would also include the quasi-Westfield effect if negative marginal pricing exists. While the input and output effects are isolated from one another in Figure 3.7 and 3.8 to facilitate their comparison, they will later be integrated in Chapter 4 to show how the firm combines all of the distortions into a determinate equilibrium solution for constrained profit-maximization.

A caveat

A word of caution is in order concerning Figure 3.8. While the isocost curve tangent to Q at point a must be steeper than the isocost curve tangent to Q' at point b or Q" at point c, it is most improbable that all three isocost curves will share the same L-axis intercept. For the sake of simplicity, this unique relationship was used in both graphs of Figure 3.8. If the MPP of K is positive and if final demand is sufficiently price-elastic, isocost C_3 could have a vertical intercept that is higher or lower than the C_2 intercept. Similarly, the intercept of C_2 could exceed that of C_3 or fall short of it. Consequently, the Hicks expansionary output effect[30] could be greater or smaller than output Q" less output Q.

---

[30] Note that the Hicks output effect (which is expansionary, given normal factors) is the desired output effect of regulation. On the other hand, the KW effect is undesirable

This word of caution should not be interpreted to mean, however, that cost employment and price consumption curves are analogous for special cases only. On the contrary, their analogous relationship can be generalized. As a point of illustration, the output effect that can separate the isocost vertical intercepts in Figure 3.8(a) is analogous to the income effect that can separate the isocost vertical intercepts in Figure 3.8(b).

AJ Distortions Versus Monopoly Restriction

The curious allegation that the input and output distortions of regulation may actually be beneficial is sometimes encountered in the literature. No less distinguished an economist than Kahn [1971, p. 59] contends that the AJ distortions ". . . could well be a more important influence for good than for poor performance." It is Kahn's argument[31] that

> ". . . [the AJ tendency] could cause production to be carried beyond the optimal point, where MC = AR . . . . On the other hand, this tendency could merely offset undesirable monopolistic restriction of output . . . . [Kahn, 1971, p. 101, n.]

---

since it can only be produced by below-MC pricing.

Although the Hicks substitution effect is undesirable as an input effect, it can certainly be desirable as an output effect, since one purpose of regulation may be to increase the output of X relative to other goods. The Hicks substitution effect becomes a detrimental output effect only when it results from below-MC pricing of good X.

[31] Kahn [1971, p. 101] actually uses this argument in reference to sales maximization and the AJ tendency, both of which he believes may ". . . lead to overinvestment and

It is rather unclear whether Kahn's reference to the AJ "tendency" is meant to include input as well as output distortions. This confusion is of little importance, however, since Kahn's reasoning can be shown to be specious regardless of which distortion is under consideration. Let it first be recognized, however, that regulation can have a beneficial, expansionary output effect. The regulatory constraint can induce the firm to equate marginal price with MC and expand production to the socially efficient level of output. Even so, this desired expansion of output, which does offset monopolistic restriction of output, must be attributed to the restraint itself and not to the distortions induced by the restraint.

Undesirable Input Effects

It should be obvious that the AJ input distortion has no socially desirable effect on output, since it raises costs and produces the AJ contractionary effect on output. This perverse influence serves to offset regulation's desired expansionary effect and to reinforce the monopolistic restrictive effect on output. Moreover, it distorts the ratio in which factors are combined, and this result could hardly

---

overproduction, with companies undertaking price reductions or promotional campaigns to produce additional sales the value of which to buyers is less than the additional costs that they entail." Although Kahn's reasoning is specious with regard to sales maximization also, it shall not be explored here because it lies outside the scope of this discussion.

be considered desirable at any level of output. As an example, even when output is suboptimal, the desired result is that more of both factors be employed (efficiently)--not that more of only one factor be employed. Finally, these undesirable results of the AJ input effect are simply worsened by the Westfield effect.

Undesirable Output Effects

It should be equally clear that the AJ expansionary output effect (here labeled the "KW effect"), which results from below-MC pricing, cannot be socially desirable even when it does offset monopolistic restriction of output. The familiar textbook argument[32] states that price reflects social value of the marginal production unit and that MC measures social cost of the marginal unit. As a result, below-MC pricing causes a net loss of social welfare on each of the units sold in this manner.

The net effect of below-MC pricing on social welfare is illustrated in Figure 3.9. If price discrimination were

---

[32] Scherer [p. 522, ft. 7] reports that MC pricing was proposed for utilities as early as 1844 by Jules Dupuit. See "On the Measurement of the Utility of Public Works," published in translation in International Economic Papers, No. 2 (London: 1952), pp. 83-110. Of course, Kahn [1971, p. 101] is familiar with this argument, acknowledging that promotional sales may have a ". . . value of which to buyers is less than the additional costs that they entail." Nonetheless, he concludes later in the text [1971, p. 107] that "as an offset to monopoly, the A-J-W distortion probably does more good than harm." As noted previously, Kahn includes Wellisz in his "A-J-W" abbreviation.

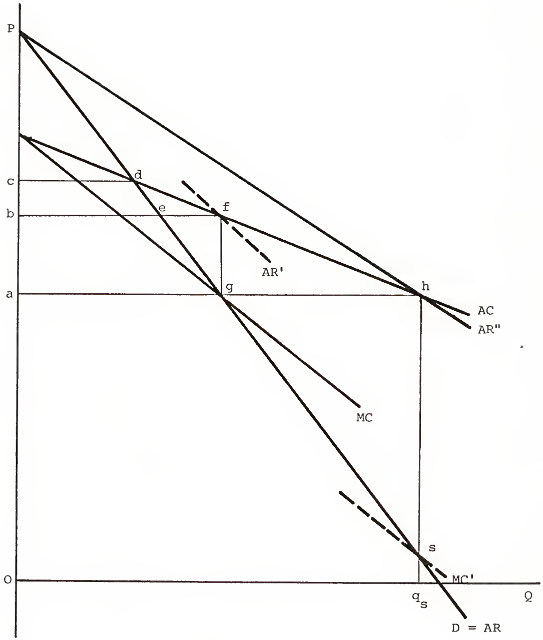


FIGURE 3.9

EFFECT OF PRICE DISCRIMINATION ON
CONSUMER AND PRODUCER SURPLUS

not employed, the constrained firm would operate at point d, where AC = AR. However, society would be better served if this firm were to use price discrimination to increase AR to the dashed position at AR'. This policy would enable the firm to operate at position g, at which point marginal price equals marginal cost. Because the cost curves are negatively sloped,[33] the firm's movement from point d to g causes it to incur a negative producer's surplus equal to area efg. Nevertheless, this deficit can be easily financed by cross-subsidies taken from the increase in consumers' surplus, the increase being equal to area bcde. Because the latter area is clearly larger than the former, it is evident that price discrimination increases social welfare when the marginal price is allowed to fall to the level of marginal cost.

This result demonstrates the desired expansionary effect of regulation, but it should not be misconstrued to mean that the AJ expansionary effect (i.e., the KW effect) is also desirable. For purposes of illustration, let it be assumed that the firm's socially efficient point of operation was initially located at position s, where marginal cost is MC' and output is q_s. Because of the AJ input effect, regulation is assumed to have raised the cost curves to MC and AC. If price discrimination is employed to move the firm from position g to the initial social optimum at point s, the increase in consumers' surplus (area abeg) will only partially offset the decrease in consumer's surplus (area ghs). Consequently, social welfare suffers a net loss when below-MC pricing is used by the firm. The AJ expansionary output effect is socially detrimental, then, even when it only serves to return production to the socially optimum level (which should be defined to be efficient only in the absence of the cost-raising input distortion). Of course, below-MC pricing is even more deleterious when it is wielded to force output beyond the socially efficient level (which could easily be the case if first-degree price discrimination were actually utilized).[34]

---

[33] Negatively sloped cost curves are used in Figure 3.9, because Kahn [1968, p. 172] believes decreasing costs to strengthen his argument. That is, long-run decreasing costs are said to increase the importance of monopoly power and decrease the importance of the AJ distortions. Since Kahn's conclusion is shown (below) to be untenable even when cost conditions are most favorable to his case, his argument cannot be salvaged by constant or increasing cost assumptions.

Kahn [1968, pp. 170-73] counters this argument by claiming that cost and demand uncertainties could render the regulatory commission powerless to force rate reductions, since regulation alone cannot force the firm to probe demand elasticity. This presents no problem in a static model, because the firm is assumed to arrive at its equilibrium position instantaneously. Within a dynamic framework, however, the firm must be willing to explore its cost and revenue curves. If the firm has a tendency to underestimate either the possibility of scale economies or the demand elasticity, the AJ effect is claimed to be beneficial to society because it moves the firm toward the competitive level of output. An examination of this contention is facilitated by use of Figure 3.9. Assuming that price discrimination is not allowed, the firm would fail to explore its demand elasticity only if it produces an output smaller than that associated with point d. However, if this were the case, AR would exceed AC and it could not be said that the commission is powerless to order a general rate reduction. Assuming that price discrimination is permitted, the firm has a natural incentive to lower marginal price and locate at position f, because this expands the rate base and increases allowable profits. If the firm is so cautious that it fails to probe section deg of the demand curve all the way to point g, it must be admitted that the commission is powerless to offset this hesitancy. Even so, it should be recognized that the AJ distortions are also powerless to effect a change in the firm's location from point d to point f. The AJ input distortion is ineffectual because it produces a contractionary influence on output (the result of cost increases). The AJ expansionary output effect is ineffectual, because the firm need not use below-MC pricing to move from point d to f. Below-AC marginal prices (that are above or equal to MC) are quite sufficient. Finally, it can be acknowledged that the AJ expansionary output effect is needed to move the firm from point f to position h (which is associated with the competitive level of output). However, operation beyond point f is clearly undesirable since social welfare is diminished, as was shown earlier. Within the static or dynamic framework, then, AJ distortions cannot be justifiably defended as an offset to the monopolistic restriction of output.

---

[34] The model in Figure 3.9 makes the rather extreme assumption that marginal cost is raised all the way from MC' to MC by the AJ input distortion. It would be more realistic to suppose the change in MC to be far less extreme and to suppose the use of price discrimination to fall short of the first-degree. Nonetheless, the unique situation envisioned in Figure 3.9 greatly simplifies the graphics without altering the basic conclusion that below-MC pricing is detrimental.

CHAPTER 4

INPUT AND OUTPUT DISTORTIONS IN THE MULTIPLE-PRODUCT FIRM

Introduction and Summary

While the previous two chapters showed how the constrained firm employs either input or output distortions, the present chapter explains how the firm optimally combines these distortions to maximize profits. Moreover, while the previous analysis was confined to a study of the single-product firm, the present analysis encompasses the more realistic case of a multi-product firm. The primary objective of this chapter, then, is the construction of a determinate model showing the profit-maximizing combination of input and output distortions in the multiple-product firm. Toward this end, a series of increasingly realistic (and therefore complex) output models and input models[1] will be derived from the Clemens and Zajac models, respectively. The relationship between these two types of models will be examined to reveal the interdependency existing between input and output distortions.

---

[1] Unlike other chapters in this thesis, the present chapter reverses the usual procedure of discussing input effects before output effects. This procedure allows the determinate model, which occupies input space, to be developed without interruption from the input models.

The Clemens model of price discrimination and the multiproduct firm will be extended to make it applicable to constrained profit-maximization. Several output models will be developed as the basic Clemens model undergoes cumulative modifications to allow standardization of cost curves, introduction of first-degree price discrimination, and adjustment for different capital intensities. Three input models, each an extension of the basic Zajac framework, will be constructed to show input distortions in a two-product firm under three alternative conditions: (1) production functions being identical, (2) production functions being different while each has fixed factor proportions, and (3) production functions being different while each has variable factor proportions. Because the last of these models is indeterminate, a more complex, fourth model is developed that yields a determinate equilibrium solution. In addition, this model will be used to identify the equilibrium levels of output for firms having preference functions other than profit-maximization.

In the previous chapter, the output distortions were called either KW or quasi-Westfield effects, depending upon whether the below-MC marginal pricing was positive or negative. Due to the introduction of a two-product firm in this chapter, output distortions will also be classified in a second way: distortions in the marginal rates of substitution of regulated products X_1 and X_2 for unregulated Y will be labelled "Type 1"; distortions in the marginal rates of substitution of the more K-intensive X_2 for the less K-intensive X_1 will be referred to as "Type 2". This distinction is highly significant, because a trade-off will be shown to exist among the input and two output distortions. Identification of the trade-off yields this very important conclusion: because a constrained multi-product firm may partially (or, indeed, completely) substitute either the Type 1 or Type 2 output distortion for the input distortion, recognition of the output inefficiencies reduces the theoretical significance of the AJ and Westfield effects.

Multi-Product Models of Output Distortion

Clemens Model of Price Discrimination

Clemens argues that a firm's use of multiple-product sales may serve essentially the same purpose as the use of single-market price discrimination. Because the latter practice is illegal, the firm may choose instead to differentiate its single product into multiple products, which can be priced according to demand elasticity to earn varying percentages of profit. In this way, the firm will tend to carry production to the point at which the least profitable product (in its product line) will be priced nearly equal to marginal cost.

To obtain this result, Clemens designs his model as follows. Market structure within each of the firm's various markets is assumed to range from monopoly to nearly perfect competition. The firm starts out as a single-product producer in a lucrative monopolistic market. Even though the firm is equating MC and MR, it finds that it is only operating at about two-thirds of plant capacity. Because the firm has mobile resources, it can take advantage of the excess capacity to produce a variety of products. The firm will look for new markets where demand still exceeds marginal cost, so that it can expand output and sales without having to lower price in the initial market. Consequently, the firm is induced to compete by invading new markets with the use of price discrimination (between market groups). With this reasoning, Clemens reaches the paradoxical conclusion that price discrimination becomes the most common form of competition.[2]

The Clemens graphical model will be demonstrated with the use of Figure 4.1, which assumes the firm to be operating in four markets.[3] The markets are invaded in the order of their profitability. The firm reaches its equilibrium position only after it has invaded every accessible market in which marginal cost is less than marginal revenue. Profit maximization occurs when the firm distributes production

---

[2] Although Clemens [p. 266] uses the term "competition," it would more appropriately be called "rivalry."

[3] Clemens [p. 267] actually employs five markets in his example. Theoretically, at least, there is no fixed limit to the number of markets that can be invaded.

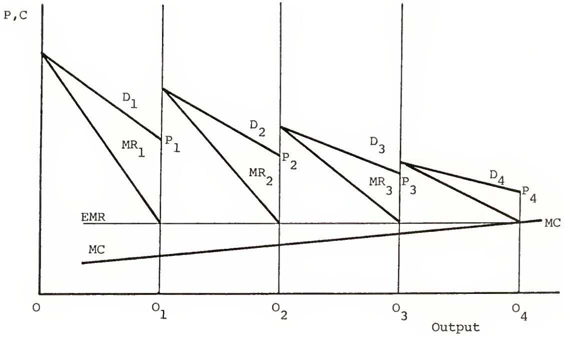


FIGURE 4.1

CLEMENS' MODEL OF PRICE DISCRIMINATION
AND THE MULTIPLE-PRODUCT FIRM

among the four markets in a way that equates marginal revenue in all markets which in turn are equated to marginal cost of production. The EMR line is drawn to indicate a level of equal marginal revenue. Each market is shown to have its own "zero output axis," from which its average and marginal revenue curves originate. The firm's demand, then, is considered to be broken down into a series of demand curves, and the EMR line determines the firm's output in each market.[4]

The MC curve is assumed to be continuous and, presumably, undisturbed by differentiation of the firm's product. The four products, then, have common costs, while no special costs can be attributed to any product. This means that the MC curve rises--not because the marginal products are more costly than the initial products--but because the plant operations are less efficient as output expands.

The four markets are ordered from left to right in accordance with their "chronological entry." It is further assumed that markets are entered in accordance to their profitability. As Clemens states, these two requirements mean that a firm will align its markets according to the following three standards:

---

[4] Clemens [p. 267] notes that if the number of accessible markets were assumed to be "certain and limited," the EMR line would need to be derived by a method used by Robinson [chaps. 15 and 16]. That is, an aggregate marginal revenue curve would be obtained by horizontally adding the various marginal revenue curves. The intersection of this aggregate MR curve with the MC curve then determines aggregate output and the EMR line's position.

> . . . (1) inversely according to the elasticity of demand in each market, (2) according to the margin of profit measured as a percentage of either standard direct costs of price and (3) according to price per standard unit of output. [Clemens, p. 268]

This means that the level of the EMR line is determined by the intersection of the MC curve (for the plant) and the MR curve (for the last market which the firm can enter profitably). The marginal market will be the market having the most elastic demand, and its price will be established at a level that nearly equals MR and MC. If demand in the marginal market is less than perfectly elastic, the possibility that a profitable market is left unserved exists. Were this the case, however, the firm could not be said to be in equilibrium.

It is assumed in Figure 4.1, then, that no fifth market exists which has a demand curve lying above the EMR line. In the event that such a market should develop, the EMR line would be re-established at a higher level, restricting production in each of the existing four markets. If the firm had been operating near capacity so that MC is rising rapidly, the new market might raise the EMR line to such an extent that the demand for an existing product is "submerged." In this event, the old product would be dropped from the product line.

Clemens credits his model with showing the dynamic means by which a firm maximizes profit while invading markets according to their profitability. He also credits the model with the ability to focus attention on the marginal market and on the ". . . spatial nature of the total market . . ." [Clemens, p. 271]. In addition, portions of the MC curve are associated with certain output blocks, an association which he believes to be consistent with the way management thinks. Finally, Clemens' most important conclusion is that the model shows the difference between price and marginal cost to be much smaller than is commonly recognized in the analysis of single-product firms.

Modified Clemens Model: the Constrained Firm

Clemens' model of the unconstrained, multiple-product firm provides a simple framework which, after three modifications, will be used to demonstrate the two different output effects produced by regulation. The first modification is the elimination of the assumption that the products share common costs. It will be assumed instead that the firm's products have different and completely independent production functions. The second adjustment will introduce the use of first-degree price discrimination within markets, in addition to the model's present use of third-degree price discrimination among markets. Finally, a third change will allow the separate production functions to have different K/L ratios, even though these ratios will still be assumed to remain fixed. An explanation of why these changes are needed and what effects they will have upon the original model follows.

Standardization of cost curves

The Clemens model assumes that the firm's products share common costs and that none has any special costs associated with it. Presumably, product differentiation is accomplished by means of nearly costless devices. For example, the several products may be differentiated by color, advertising claims, or package design. The assumption that all costs are common does not diminish the model's potential for showing one type of output distortion.[5] However, the modified model that demonstrates output distortions will later be related to the model demonstrating input distortions. The latter of these two models will be greatly simplified by the assumption that production functions are independent of each other. Consistency will be maintained, then, if the eventual need for this assumption is anticipated now. Consequently, the first modification to be made of the Clemens' model is the assumption that all costs of the separate products are special rather than common.[6]

When the production functions are assumed to be different from each other, the marginal cost curves must be standardized before the markets can be aligned from left to right in the modified model. The need for this adjustment is apparent in Figure 4.2(a), which shows the level of marginal costs for products 1 and 3 to be double that for products 2 and 4. For added simplicity, average and marginal costs are here assumed to be constant for each product, though this assumption is not essential. Delineation of the output axes is the same for each of the four markets in Figure 4.2(a). This means that if one unit represents a pint of output in market 1, one unit equals a pint in the other three markets also.

---

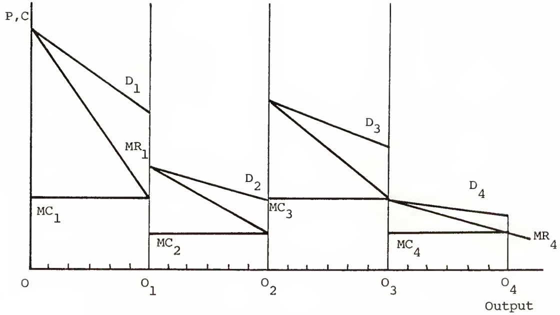
[5] The common-cost assumption does, however, preclude the demonstration of a second type of output distortion (to be identified later in this chapter).

[6] Joint costs also must be ruled out, but Clemens [p. 266] specifically excludes these himself.

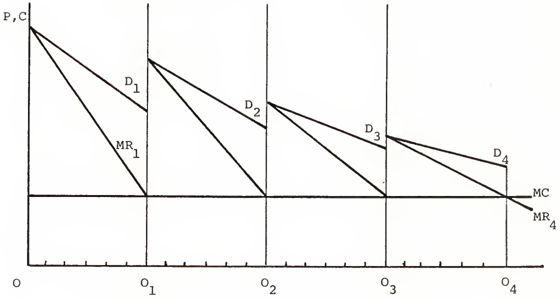
In this example, standardization of the MC curves is easily accomplished by defining the units in markets 1 and 3 to be pints and the units in markets 2 and 4 to be quarts. The MC of a quart is twice that of a pint, so the MC curves for products 2 and 4 are raised into alignment with the MC curves for the other two markets. Similarly, a quart will command the price of two pints, so every point on demand curves D_2 and D_4 is raised to a level that is twice its original height. This adjustment is demonstrated in part (b) of Figure 4.2.[7] Because MC is now the same for each market it is apparent in part (b) that the four markets are correctly alligned in order of profitability to the firm. It is also clear that the equilibrium levels of output were correctly

---

[7] Note that the output axes of markets 2 and 4 have been recalibrated, making the length of a unit in part (b) twice what it was in part (a) for these two markets. This recalibration is not essential, since it only serves to maintain the horizontal distances between the vertical axes in part (b) as they were in part (a).



(a) Homogeneous Delineation of Output Axes



(b) Heterogeneous Delineation of Output Axes

FIGURE 4.2

STANDARDIZATION OF COST CURVES FOR THE MULTIPLE-PRODUCT FIRM.

represented in part (a) of the figure. That is, before the adjustment the firm was electing to sell six pints of product 2 at a price that was double average cost. After the adjustment, equilibrium is undisturbed, since output is still six pints (i.e., three quarts) and price is still double average cost. It follows, then, that standardization of the cost curves enables the model to encompass disparate production functions.[8]

Introduction of first-degree price discrimination

It was shown in Chapter 3 that the constrained firm has a profit incentive to use first-degree price discrimination when producing only a single product. Likewise, the constrained, multiple-product firm will maximize profits only when it uses first-degree price discrimination within each of its separate product markets. In this way, the firm uses third-degree discrimination to break up aggregate demand into markets small enough that first-degree discrimination can be employed within each. As a result, consumer surplus is fully extracted in one way or the other.

When it is assumed that the K/L ratio is the same for the production of each product, the introduction of first-degree price discrimination has a straightforward effect on the modified model. Each of the four average revenue curves in Figure 4.2(b) will swing outward until its slope is reduced by half.[9] Similarly, each MR curve swings outward until it coincides with the corresponding demand curve. The effect on the model is illustrated in Figure 4.3. Within each market, MR now equals demand. In market 1, for example, the former position of the MR curve is indicated by the dashed line descending below its new position at MR_1.

---

[8] The standardization is actually suggested by Clemens, who writes that "the shape and level of the marginal cost curve would also be influenced by the composition of the standard units of output and the standard direct costs assigned to each of the products" [Clemens, p. 270 n.].

The equilibrium levels of output, even though all have expanded, are still determined by the equal marginal revenue (EMR) line. Note, however, that MR no longer need be equal to MC. The regulated firm in Figure 4.3 is earning such excessive profits (i.e., the striped areas) in markets 1 and 2 that it is encouraged to expand output into markets 3 and 4. This penetration of new markets enables the firm to increase its capital rate base, which justifies greater profits. Even though the firm breaks even in market 3, its loss (i.e., the dotted area) in market 4 serves to reduce earnings to an acceptable level. To determine the equilibrium levels of output, the EMR line is raised or lowered until the aggregate profit area exceeds the aggregate loss area by an amount commensurate with the allowed rate of return.[10] The modified

---

[9] The effect of first-degree price discrimination on the AR curve was explained earlier in Figure 3.4.

[10] After developing this graphical analysis in which the profit areas are offset by the loss areas, I have found that Bailey and White [pp. 16-18] use a similar technique. Their

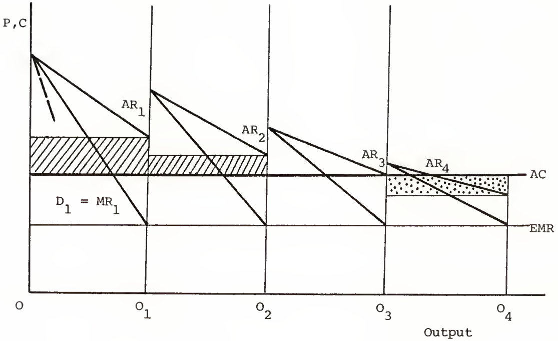


FIGURE 4.3

THE CONSTRAINED, MULTIPLE-PRODUCT FIRM

model in Figure 4.3 does not specify the effect of output changes on the rate base. However, a determinate model which does this will be offered later in this chapter when the input and output models are integrated.

Adjustment for different capital intensities

Capital intensity coefficient.--Because the products have independent production functions, it seems reasonable to assume that capital intensity varies among the several products, even though the K/L ratio is fixed[11] for any one product. This assumption requires that the Clemens model be modified in yet a third way, which is illustrated for one of the markets in Figure 4.4. The K/L ratio for each product can be indicated by a capital intensity coefficient. The most capital intensive product is given a coefficient of 1. The K/L ratio of this product is the standard against which the ratios of all other products are compared. Thus, a product whose K/L ratio is only half that of the reference product receives a coefficient of one-half. Likewise, a product which is produced with labor alone has a coefficient of $1/\infty = 0$.

---

framework is a peak-load pricing model, however, and they assume the cost curve to include a permitted return on capital. The cost curves in Figure 4.3, on the other hand, necessarily exclude the return on capital.

[11] The model is greatly complicated when the capital intensity of each product is allowed to vary. Consideration of this more interesting situation, then, is postponed to the end of this chapter.

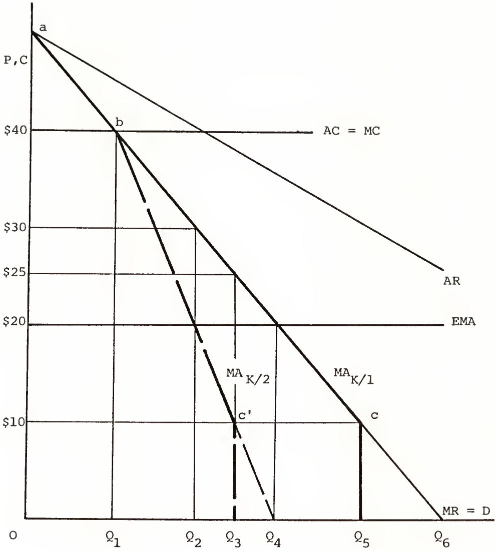


FIGURE 4.4

DERIVATION OF THE MARGINAL ASSESSMENT CURVE

If the equilibrium level of MR is \$20 for all products in the line, and if the K coefficient for the product equals 1, then the equilibrium output of the product in Figure 4.4 is Q_4. At this level of output, AC and MC (both equal to \$40) exceed MR by \$20. Because the firm's marginal loss is \$20, the firm is only having to spend this amount to acquire the capital needed to produce that last unit of output. The market cost of this amount of capital must be greater than \$20, or the firm would have simply elected to acquire nonproductive capital instead of expanding output.

On the other hand, the product depicted in Figure 4.4 might be produced with a capital intensity coefficient of only one-half. As long as MR exceeds MC, output will not be affected by the product's reduced capital intensity. Whenever additional sales of a product increase total profits, more will be sold whether production requires the use of capital or not. When MR falls below MC, however, the firm will increase output only if it is willing to regard the marginal loss (i.e., MC less MR) as a necessary expense for a larger capital rate base. Because the product in question only uses half as much capital as the reference product, the marginal loss that the firm is willing to sustain for greater output is only half as large, also. Accordingly, the firm is only willing to lose up to \$10 per unit of output. The equilibrium output, then, is Q_2, at which level MC (still equal to \$30) exceeds MR (now equal to \$30) by \$10.

Marginal assessment curve.--The equilibrium level of output of the K/2 product is no longer determined by the marginal revenue curve abcQ_6. Instead, it is determined by a marginal assessment (MA) curve abc'Q_4, labeled $MA_{K/2}$. The MR and $MA_{K/2}$ curves coincide between points a and b, because output is unaffected by capital intensity as long as MR exceeds MC. Below point b, however, the slope of the $MA_{K/2}$ curve is twice that of the MR curve, reflecting the fact that the K/2 product is only worth half as much as the reference product as a rate base contributor.[12] Similarly, the K/3 product's MA curve would be three times as steep as that for the reference product ($MA_{K/1}$) below point b. The equilibrium levels of output in all markets are now determined by an equal marginal assessment line (EMA) instead of Clemens' EMR line. This insures that a $1 loss in one market will add just as much capital to the rate base as a $1 loss in any other market. The EMR line is lowered until, as before, the profit area exceeds the loss area by an acceptable amount.

---

[12] Averch and Johnson [pp. 1058-1059] state that the constrained firm that is making a profit in each market will equate MR and MC within each market. More important, A and J recognize that when the firm is sustaining losses in at least one market it will not simply equate MR and MC to determine market prices. They argue, instead, that the firm will seek ". . . to equate the marginal value product of capital in both markets. . ." [Averch and Johnson, p. 1059].

If their term, "marginal value product," is taken to mean "value of marginal product," their statement is incorrect. Because neither market in their example is perfectly competitive, the firm would strive to equate the marginal revenue products--not the values of marginal products--of capital.

Even if "marginal value product" is interpreted to mean "marginal revenue product," the AJ statement would be correct only for products that are all produced with the same capital intensity level. Their analysis fails to specify whether the K/L ratios of the two products are identical or not.

Consideration of nonproductive capital.--As the EMA line is lowered, a level will be reached eventually where the firm will find it more profitable to expand its rate base through the acquisition of nonproductive capital than through output expansion. The MA curve must be altered if it is to accommodate this situation. The effect of the adjustment is most easily explained (within the context of Figure 4.4) when the following assumptions are made: the market price of capital is $10 and the supplies of both factors (K and L) are perfectly elastic. The price of nonproductive capital (K*) equals that of K. Finally, 3 units of K are needed (when efficiently combined with L) to produce one unit of the reference product or two units of the K/2 product.

In Figure 4.4, it is demonstrated that the firm will not produce an output greater than Q_5 of the reference product, no matter how low the EMA line is drawn. The MA curve for the reference product coincides with the MR curve throughout its entire length. Since at output Q_5 marginal loss equals $30 (i.e., $40 less $10) and the marginal addition to capital stock (as a result of a one unit increase in output) is 3 units of K, the marginal "expense" (i.e., loss) for K by means of output expansion is $10 per unit of K. At this level of output, then, it would be no more expensive for the firm to expand its rate base by acquiring nonproductive K*, which commands a market price of $10. If output were to expand beyond Q_5, the firm's marginal "expense" of K would exceed the marginal expense of K*, so output of the reference product will clearly be restricted to Q_5 by the market cost of capital. It follows that the MA curve for the reference product is curve abcQ_5, which becomes perfectly price-inelastic at output Q_5.

In the same way, the MA curve for the K/2 product can be shown to be perfectly inelastic when the product's MA is $10 or less. At output Q_3, the firm's marginal loss is $15 (i.e., $40 less $25). The marginal unit of output requires 1.5 units of K. At this level of output, then, the marginal "expense" on K equals $10, which would also be equal to the marginal expense of K* (if the firm were also purchasing nonproductive capital). Consequently, the MA curve for the K/2 product can be represented by curve abc'Q_3. Note that the MA curves for both products (and all other products in the line) become vertical at the same level of marginal assessment, which equals $10 in this example.[13]

Equilibrium levels of output.--When capital intensity is allowed to differ among products, the equilibrium levels of output (and therefore, prices) are determined by the EMA line which is lowered until total profit reaches a permissible level. The marginal assessment of output will be the same in all of the firm's markets. This situation is illustrated in Figure 4.5, in which it is assumed that MR = D within each of the four markets. Output in the first market is produced with labor alone. Since the MA_1 function below MC is obtained by dividing the MR_1 function by the K intensity coefficient, the slope of the MA_1 curve for the K/∞ product is infinitely steep below the MC line (i.e., $MR_1/0 = \infty$). In market 1, then, MA_1 coincides with the MR_1 curve above MC but is vertical below MC. The EMA line intersects this vertical portion of the MA_1 curve at an output equal to the distance between the 0 and 0_1 axes. The average price being charged for product 1 is indicated by the AR_1 curve, and profit is represented by the striped area between AR_1 and the AC curve.

---

[13] To simplify Figure 4.4, the level at which the MA curves become vertical was designed to equal the price of capital. These two values need not be equal under normal circumstances.

Because product 2 is the reference good, the MR_2 and MA_2 curves are identical. The equilibrium levels of output and price, then, are determined here in the same way that they were resolved in Figure 4.3, except that the EMA line is used in place of the EMR line. As before, profits are positive and are indicated by the striped area. In market 3, however, the MA_3 curve below MC has a slope that is twice as steep as the MR_3 curve, since the K intensity coefficient is one-half. The MA_3 curve intersects with the EMA line above 0_3, and the firm breaks even in this market. Nevertheless, this break-even solution should not be interpreted to mean that total profits are not increased by market 3. Instead, it

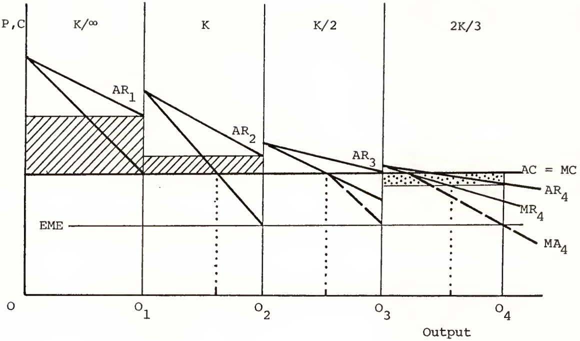


FIGURE 4.5

EQUILIBRIUM SOLUTION WITH FIRST-DEGREE PRICE DISCRIMINATION AND DIFFERENT CAPITAL INTENSITIES simply indicates that AR_3 equals AC, a situation which would still allow total profits to be increased by market 3's contribution to the rate base. Finally, in market 4 the portion of the MA_4 curve below MC has a slope equal to 3/2 that of the MR_4 curve. The firm's loss in this marginal market is represented by the dotted area.

It is only assumed in Figure 4.5 that the EMA line is at the equilibrium level, so that the profit area exceeds the loss area by the maximum amount that can be justified by the rate base. This assumption is necessary because Figure 4.5 cannot be itself provide an equilibrium solution, since it fails to disclose the effect of output changes on capital stock. The introduction of a static, determinate, equilibrium solution must await the development of the Zajac model showing input distortions, which follows in the remaining part of this chapter. Equilibrium will be determined, then, when the relationship between the input and output models is ascertained.

Multi-Product Models of Input Distortion

In Chapter 2 the Zajac model was modified to demonstrate the effect of rate base padding. This model will be modified here to show the input distortions that result from regulation of a multi-product firm. In increasing order of complexity, three different situations will be examined. The first case identifies output distortions when the firm's production functions are identical, the second when these functions are different from each other but fixed, and the third when these functions are different and variable. Each of the three input-distortion models will be related back to the output-distortion model shown in Figure 4.5. The equilibrium solutions for the first two static models are determinate, but the solution for the third remains indeterminate until the framework is extended into a fourth model at the end of the chapter.

Input Distortions with Identical Production Functions

The input-distortion model of the multi-product firm is simplest when the firm's products are assumed to have identical production functions. For even greater simplicity, the K and L inputs are assumed to be used in the production of only two products, X_1 and X_2. The resulting elementary model is represented in Figure 4.6. The profit hill (not shown) is derived from cost and revenue data provided by two markets similar to markets 2 and 4 in Figure 4.3. The efficient path in Figure 4.6 is a straight line, since average costs were earlier assumed to be constant (in the output model). As the EMR line is lowered, its intersection with the two MR curves in the output model will identify the profit-maximizing output combinations of X_1 and X_2 at each EMR level. When the EMR line is very high, for example, the firm may choose to produce X_1 only. Conversely, when the EMR is low it is conceivable

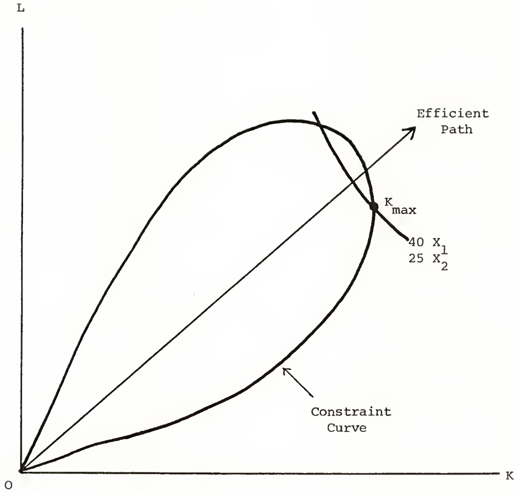


FIGURE 4.6

EQUILIBRIUM WHEN PRODUCTION FUNCTIONS
ARE IDENTICAL

that X_2 would be produced in a greater quantity than X_1. The unique series of output combinations, which are specified by the EMR line and MR curves in the output model, determine the isoquant values in the input model (i.e., Figure 4.6). In this example, isoquants very near the point of origin possess positive X_1 values only. Higher isoquants have positive output levels for both X_1 and X_2. The two-product isoquant shown in Figure 4.6 represents an output level of 40 units of X_1 and 25 units of X_2. Although they are not shown, a lower isoquant curve might represent 38 X_1 and 24 X_2 units, while a higher curve could depict 42 X_1 and 26 X_2 units. Once the isoquant values and the corresponding product prices are derived from the output model, the profit hill and constraint curve can be constructed in the input model of Figure 4.6. The K_{max} point of equilibrium indicates that the firm will maximize profits by selling 40 X_1 and 25 X_2 units. Since the production functions for both products are identical, and since the since these functions have been assumed to be linearly homogeneous,[14] the two-product isoquant in Figure 4.6 is identical to the single-product isoquant depicting 65 units of either product. The input distortion of K for L is measured as the slope and distance between the isoquant's point of intersection with the constraint curve (at K_{max}) and its point of intersection with the efficient path.

---

[14] If the production functions were not assumed to be linearly homogeneous, it would be necessary to assume that both products are manufactured together as one product. Product differentiation would be accomplished in some costless manner (i.e., color or advertising differences). This assumption would insure that both products would be affected by the same economies or diseconomies of scale.

Input Distortions with Different, Fixed Production Functions

When the two products have fixed production functions that are different from each other, no input distortions will be produced (even though output distortions may well be present). This situation, which is only a little more complicated than the preceeding case, is shown in Figure 4.7. As before, the values of the two-product isoquant curve are derived from the output model. Also as before, the equilibrium output levels at K_{max} are given to be 40 X_1 and 25 X_2 units. Because the production functions are linearly homogeneous, a simple geometric technique can be used to locate the equilibrium positions of the single-product isoquants. A dotted line intersecting K_{max} is drawn parallel to the X_2 expansion path. The point of intersection of this line with the X_1 expansion path identifies the corner of the X_1 isoquant having a value of 40 units. The X_2 isoquant curve having the equilibrium value of 25 units can be identified in a similar way. Of course, when the K and L values for each of the single-product isoquants are aggregated they will equal the K and L values associated with K_{max} on the two-product isoquant curve. It is equally obvious that the input distortion will equal zero, since the two expansion paths are also the efficiency paths.

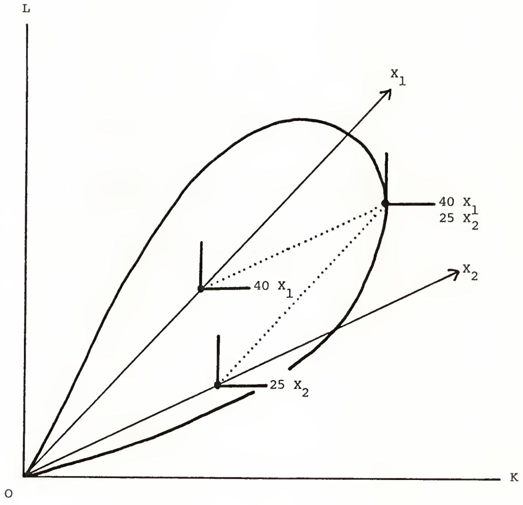


FIGURE 4.7

EQUILIBRIUM WHEN PRODUCTION FUNCTIONS ARE FIXED AND DIFFERENT

Input Distortions with Different, Variable Production Functions

The input-distortion model is greatly complicated when the separate production functions are both assumed to be variable. In Figure 4.8, the two-product isoquant is derived from the two variable single-product isoquants. When factor price lines having identical slopes are drawn tangent to the three isoquants shown, the aggregate K and L values for the two points of tangency on the single-product isoquants will equal the K and L values associated with the point of tangency on the two-product isoquant curve. The slope of the line drawn tangent to the aggregate isoquant at K_{max} is the (negative) ratio of the firm's effective cost of K to the market cost of L. Of course, the efficient point of operation on the aggregate isoquant curve is designated by the curve's tangency with an isocost line that compares market costs for both factors. The two-product isoquant segment between the efficient point and K_{max} measures the input distortion of K for L. Note that the equilibrium output levels of X_1 and X_2 are here only assumed to be 40 units and 25 units, respectively. Unlike the previous two input models, this model is not sufficiently detailed to yield a determinate equilibrium solution--even when it is related back to the output-distortion model. The problem is one of interdependency. The output model is said to determine output combinations in the input model. Nevertheless, the selection of output combinations

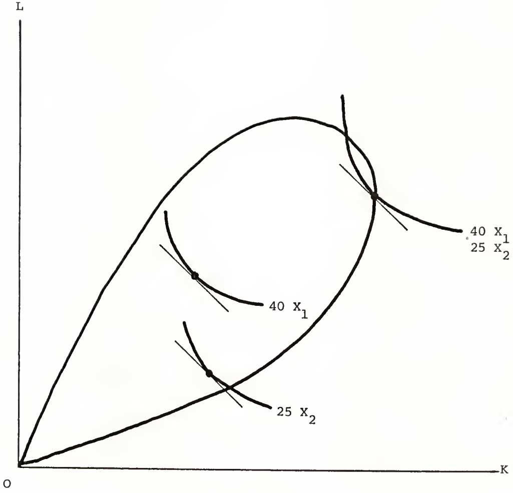


FIGURE 4.8

EQUILIBRIUM WHEN PRODUCTION FUNCTIONS
ARE VARIABLE AND DIFFERENT

also affects the relative capital intensities of each product in the input model. This change simultaneously repositions the MA curves in the output model, which in turn disturbs the original output combination. This theoretical framework can be made determinate, however, when the model is extended appropriately. It is to this end that the remaining part of this chapter is devoted.

Multi-Product Model of Input and Output Distortions, Given Variable Production Functions

Optimal Combination of Input and Output Distortions

Figure 4.9 shows how the constrained, two-product firm will employ input and output distortions in combination to maximize allowable profits, which requires that the K_{max} position be moved as far right as possible. The unconstrained monopoly which does not use price discrimination will operate at point M. When the firm is regulated and, at the same time, permitted to use just enough price discrimination to equate MP and MC, the profit hill (not shown) will expand and constraint curve CC will be established.[15] The unconstrained profit-maximizing point, given this same degree of

---

[15] Presumably, the unregulated firm will not price-discriminate for fear of violating section 2(a) of the Robinson-Patman Act. The regulated firm, on the other hand, may use a high degree of price discrimination, since this does not violate the simple rate-of-return constraint and since the FTC may not be so likely to interfere with the regulatory agency.

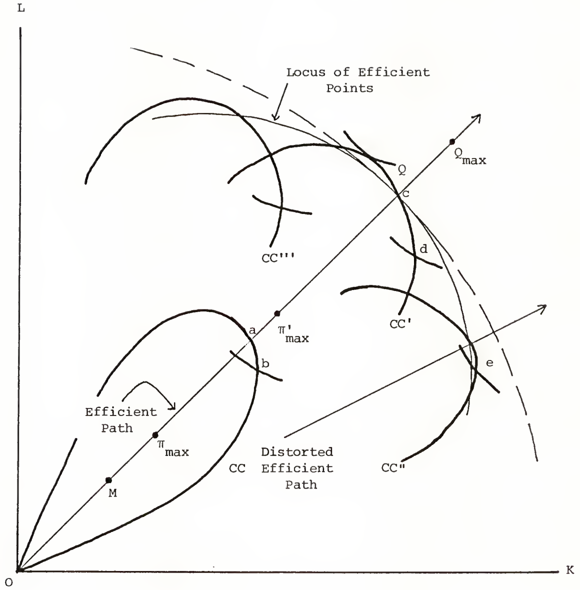


FIGURE 4.9

COMBINED EFFECT OF INPUT AND OUTPUT DISTORTIONS ON THE CONSTRAINED EQUILIBRIUM POSITION price discrimination, is π_{max}. The socially optimum levels of output would be produced if the firm chose to operate at position a, the point of intersection of the efficient path with curve CC. Outputs would be optimal at that point because MP equals MC in each of the two markets.[16]

Effect of distortions used separately

Input distortion only.--By distorting factor combination alone, the firm can increase allowable profits and move to K_{max} at position b, the point of intersection of the two-product isoquant and curve CC. This solution is an extended version of Zajac's graphical model,[17] since the isoquant is multi-product and has been enlarged beyond Zajac's constraint curve through the use of second- or third-degree price discrimination.[18] However, if the firm were going to make a choice between using input or output distortions, it would certainly favor output distortions in this situation.

Output distortion only.--By employing below-MC pricing instead of the AJ or Westfield effects, the firm can move outward along the efficient path and away from efficient point a on curve CC. In the same way that the use of third-degree price discrimination expanded Zajac's constraint curve (not shown) outward to position CC, the use of first-degree price discrimination within the two markets will expand curve CC outward to position CC'. The firm's new point of operation is position c, located at the intersection of curve CC' and the efficient path. Again, output of each product at point c is determined by the EMR line in a two-market output model (as was explained in Figure 4.3). It can be observed in Figure 4.9 that the efficient point c on curve CC' uses about twice as much capital as the efficient position a on curve CC. This means that the firm has approximately doubled its total profits by using below-MC pricing, since its rate of return is the same on curve CC' as it was on CC.

---

[16] Output in each market would be analogous to output q_1 in Figure 3.4 or point c in Figure 3.5(a).

[17] Zajac's basic model was earlier introduced as Figure 2.2(b).

[18] Because Zajac's model includes neither price discrimination nor government subsidization, the decreasing-cost firm will not produce (in the long run) beyond the level of output at which P equals AC. In contrast, the firm that is restrained by curve CC in Figure 4.9 can choose to operate at the socially efficient levels of output where MP equals MC. Of course, the constrained profit-maximizer would not make such a choice.

Effect of distortions used collectively

Even though profits have been almost doubled by use of the output distortion alone, they can increase still further if the firm elects to distort factor combinations also. Such an action would relocate the firm's operating position from c to K'_{max} at point d.[19] Even at this new position, however, profits have not yet been maximized. Although the firm is now using both input and output distortions, it is employing only one of the two types of output distortions available. In addition to distorting the marginal rates of substitution of X_1 and X_2 for unregulated product Y (a Type 1 output distortion), the firm can also distort the marginal rates of substitution of X_1 for X_2 (a Type 2 output distortion). This second type of output distortion becomes increasingly attractive to the firm as the capital intensities of the two regulated products become ever more divergent.[20] In other words, the more capital-intensive X_2 will be promoted over the less capital-intensive X_1 by means of relative price changes.

---

[19] The product output levels at point d, like those for efficient point c, are determined in a manner illustrated earlier in Figure 4.3.

As the firm encourages consumers to substitute X_2 for X_1, constraint curve CC' swings downward to the right. If the size and shape of the profit hill were not altered by this relocation, the movement of efficient point c (assuming that CC' is simply swung about the point of origin) would trace out the dashed arc shown in Figure 4.8. Of course, the profit hill will actually shrink as one product is promoted at the expense of the other, because the firm is no longer efficiently equating the MRP of capital in both markets. This means that constraint curve CC' will contract whether it is moved right toward position CC" or left toward position CC''' . Consequently, as the constraint curve is moved about the point of origin, a locus of efficient points will be traced out which is everywhere inside the arc except for its point of tangency with the arc at position c. Note that the locus is efficient only with regard to factor combinations--not output combinations--since first-degree price discrimination is employed everywhere along the locus.

---

[20] For simplicity, the model excludes the possibility of factor reversal in the production of X_1 and X_2. The feasibility of factor reversal was suggested to me by David Kaserman.

The firm finally reaches the point of profit-maximization at position e when the constraint curve is moved to position CC". This determinate equilibrium point is reached by the firm through a trial and error process that is illustrated in Figure 4.10. The movement of the constraint curve about the point of origin traces out a locus of K_{max} points. As the constraint curve moves rightward from position CC' to CC", the capital stock increases continually. When K_{max} reaches its global maximum at K^*_{max}, however, the rightward swing of the constraint curve is exactly offset by its leftward contraction. At K^*_{max}, then, the locus of K_{max} points attains the highest level of capital stock that the profit-maximizing firm desires. Further use of the Type 2 output distortion would only cause the locus to bend back toward the point of origin. The equilibrium levels of output are specified by the two-product isoquant intersecting K^*_{max}. This isoquant relates back to the single-product isoquants for 40 X_1 and 25 X_2 in the manner that was described in Figure 4.8.

Relationship Among Input and Output Distortions

Returning to Figure 4.9, it can be observed that the

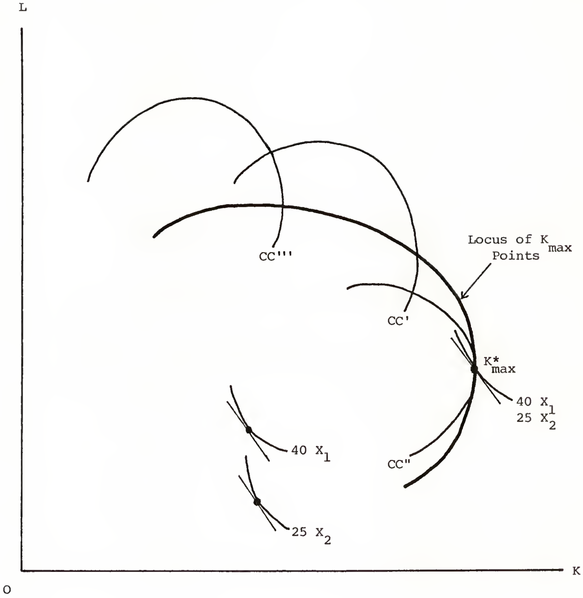


FIGURE 4.10

DETERMINATE EQUILIBRIUM SOLUTION FOR THE CONSTRAINED FIRM USING BOTH OUTPUT AND INPUT DISTORTIONS input and output distortions are all positively related to the severity of the constraint. When the rate-of-return ceiling is so high that it is ineffective, the firm operates at point M (assuming that price discrimination is successfully eliminated by the Federal Trade Commission) and no distortions are present. When the constraint becomes restrictive, the constraint curve appears, expanding continually outward toward the base of the profit hill as the allowed rate of return (f) approaches the market cost of financial capital (i). As the constraint plane is lowered, then, all three distortions are aggravated simultaneously.

On the other hand, an inverse relationship exists among the three distortion types at any given level of restraint (assuming, again, that f exceeds i but is less than the unrestrained rate of return). The effective constraint encourages use of the distortions since, on the one hand, it creates a capital premium and, on the other hand, it creates a pool of excess earnings (a portion of which can be used to finance the profit-enhancing distortions). The more of these excess funds that is used to finance one distortion type, the less that is available for the remaining two. This trade-off is quite evident in Figure 4.9. For example, the firm's movement from point c to d required that the Type 1 output distortion be partially offset by the input inefficiency. Consequently, output was reduced from its higher level at point c to a lower level at point d as the input distortion was introduced. Likewise, when the firm moved from point d to e, it partially replaced its use of the first two distortions with the Type 2 output distortion. Total sales are less at point e than they are at d. Moreover, the input distortion is also smaller, since the isoquant segment intersecting point e and the distorted efficient path is shorter than the isoquant segment intersecting point d and the efficient path.[21]

From this analysis of the trade-off comes a very important conclusion: in a regulated firm selling products that differ greatly in capital intensity, the AJ and Westfield effects may be almost entirely replaced by the Type 2 output distortion. This result may partially explain why attempts to empirically measure the input distortion have been less than successful.[22] Of course, the data problems inherent in

---

[21] Note that isoquant Q and the isoquants shown intersecting points b and d in Figure 4.9 are all part of the same two-product isoquant map. However, the two-product isoquant shown intersecting point e is part of a different isoquant map. It is one in a series of more-steeply drawn isoquants, since it has a larger component of X_2 output (which is shown to have a steeper single-product isoquant than X_2 in Figure 4.10). Similarly, the isoquant shown intersecting CC''' has a relatively shallow slope, since its X_2 component is relatively smaller.

[22] For example, in Johnson's [pp. 92-94] reassessment of the AJ effect, he cites three "bits" of empirical evidence that the effect is real. These bits can only be described as meager.

With regard to the Westfield effect, Emery [pp. 329-36] has concluded that even though the empirical evidence is mixed it does not permit the acceptance of Westfield's hypothesis. Similar beliefs are expressed by Bailey [pp. 192-95], Corey [pp. 358-73] and Kendrick [pp. 136-67].

One of the three examples of the AJ effect offered by such studies are well known. The point to be recognized here, however, (and the point to be stressed) is that the AJ effect is not simply less important empirically than is commonly believed--but it is also less important theoretically, since the firm need not rely on this distortion alone.

Alternative Managerial Preference Functions

The managers of the restrained firm may select some objective other than profit-maximization. Given different maximands, the firm's equilibrium point for each can be identified in Figure 4.9. The unregulated, profit-maximizing monopolist which is not permitted to price discriminate will operate at point M. If it were allowed to use first-degree price discrimination, the firm would locate at π'_{max}. The employment of third-degree price discrimination, on the other hand, would place π_{max} between points M and π'_{max} on the efficient path. These positions are also the alternative equilibrium points of operation (depending upon the degree of price

---

Johnson [p. 94] uses the interesting argument that AT&T may favor cable use over satellite use because its partial ownership of a cable will enlarge its rate base. An opposing argument could be made that telecommunications by cable is a different product from telecommunications by satellite. The former service may be distinguished by its greater reliability and resistance to interception by eavesdroppers. Johnson's example, then, may better support the existence of the Type 2 output distortion than that of the input distortion. In any event, it shows that these two concepts (like other classificatory devices) are separated by a grey area instead of a black line.

discrimination) for Williamson's expense-preference firm. Operation at the profit-maximizing levels of output ensures maximization of funds to be used for fringe-benefits and other preferred expenses.

The rate-of-return maximizer-cost minimizer who elects to produce the socially optimum levels of output will operate at point a. Alternatively, the Kafoglis-type manager (whose primary and secondary goals are rate-of-return- and output-maximization, respectively) will operate at the point of tangency of isoquant Q with curve CC'. Finally, the simple output maximizer (who intends to survive the long run) will reach equilibrium at Q_{max}, the point of intersection of the efficient path with the base of the profit hill (i.e., that profit hill which is circumscribed by curve CC' on the constraint plane).[23]

---

[23] In addition, the equilibrium position of Baumol's revenue-maximizer could be identified if a total revenue hill were constructed in place of the profit hill.

CHAPTER 5

CONCLUSION

It will now be appropriate to assess the success with which the three objectives (specified at the outset of this work) were accomplished. The first goal was to show that output distortions are at least as important theoretically as input distortions. Since two types of output distortions have been identified, and since each is as potentially valuable to the firm as the input distortion, it might be argued that the output distortions together pose a more serious threat to social welfare than does the input inefficiency alone. As appealing as this argument might seem, the relative values of the three distortions will have to be decided empirically, not theoretically.

It can be said here, however, that the input distortion becomes increasingly influential (relative to the size of the firm) with increases in the firm's elasticity of factor substitution. On the other hand, the importance of the Type 2 output distortion is determined by the amount of divergence between the K/L ratios of the firm's products. Consequently, the greater the divergence in capital intensities the more potent is the Type 2 distortion relative to the input distortion.

To cite an example, the Type 2 output distortion may be found to significantly reduce the amount of labor-intensive services (relative to capital-intensive services) offered by the telephone industry. The price of a person-to-person call (relative to the price of a direct-dial call) might be substantially higher due to the presence of this distortion type. Similarly, the electric utility industry may favor the more capital-intensive industrial sales over residential sales, since the latter may have a higher component of (relatively labor-intensive) customer service charges. If, on the other hand, the difference in capital-intensities between residential and industrial services is small, the Type 2 output distortion will be overshadowed by the input inefficiency.

The theory also makes it clear that the effectiveness of the Type 1 output distortion is positively related to the price cross-elasticity of demand between the regulated products of the firm and the unregulated substitutes. Below-MC pricing by the electric utility industry, for example, could enable it to compete unfairly with the unregulated coal industry. An output distortion that favors the industrial customers, whose price cross-elasticity of demand between electricity and coal is relatively high, could be subsidized by the residential and commercial customers, whose cross-elasticity of demand is relatively low.

The three-part classification of distortions as input, Type 1 output, and Type 2 output is descriptive of the influence that they have on the productive or allocative process. On the other hand, a four-part classification was also used, which subdividies the distortions into the AJ, Westfield, KW, and quasi-Westfield categories. This classification is suggestive of the causes of the inefficiencies. The AJ and Westfield effects, for example, occur because a capital premium is created by the constraint. The KW and quasi-Westfield effects, on the other hand, result from consumer cross-subsidization, which in turn is encouraged by the presence of the capital premium. The input distortions, then, are direct responses to the constraint while the output distortions are indirect (i.e., increasing capital stock through output expansion). It was this four-part classification that satisfied the second objective: to demonstrate that input and output distortions are analogous to one another. Accordingly, the AJ and Westfield effects were shown to be analogous to the KW and quasi-Westfield effects, respectively.

The recognition of this symmetrical relationship led, in turn, to the construction of a determinate model explaining the manner in which distortions can be optimally combined to achieve constrained profit-maximization. Such a model was the third objective of this work. An examination of the relationship between input and output distortions has long been needed, since the rational firm will not employ each independently of the other. Instead, the distortions are used together in constellation, enabling profits to be increased beyond the local maximum, K_{max}, to a global maximum, K^*_{max}.

It has been shown elsewhere that a tightening of the rate-of-return constraint will lead to greater input distortions. It has been argued here, along with the introduction of an analogous relationship, that the output distortions will be similarly exacerbated. Restriction of the constraint (as long as the ceiling rate of return exceeds the market cost of capital) produces a pool of excess earnings, a portion of which can be used to subsidize socially undesirable additions to capital stock. The greater the amount used to finance the input distortion, the lesser the amount available for output distortions. Likewise, the greater the revenue loss due to a Type 1 output distortion, the smaller is the revenue loss that can be tolerated from a Type 2 output distortion. A trade-off exists, then, among the three distortions.

Recognition of this trade-off is essential if the simple rate-of-return constraint is to be amended in some form to promote social efficiency. The emphasis that has been placed on input inefficiency in the literature may have led regulators to become preoccupied with that distortion type. If an amendment is too narrowly designed, discouraging the firm's use of the AJ effect only, the resulting gain in input efficiency may be almost wholly offset by the simultaneous frustration of the KW or quasi-Westfield effects, which would aggravate the Type 1 and Type 2 output distortions. In this way, a restraint which effectively counters the input distortion can produce the strange result of making a reality out of the theoretical concept of negative marginal pricing. Lest this conclusion seem far-fetched, it should be recalled that negative marginal prices are rejected (by firms which could profitably use them) only because the input distortion is more profitable yet. If the latter option alone were voided, the quasi-Westfield effect could become a frequently employed instrument for achieving constrained profit-maximization.

It seems unreasonable to suppose that all of the distortions can be simultaneously eradicated unless the regulatory agency has such complete cost and revenue information and such extensive authority that it can accurately specify the desired rate structure and output levels. However, such a situation should not be mistaken for improved government regulation. It is, instead, complete government control, which has attendant inefficiencies of its own. The increased recognition of output distortions, then, does not necessarily call for increased use of regulation. On the contrary, since regulation may be far more costly than the recognition of input distortions alone would imply, economists should place less emphasis on regulating market performance and more on changing market structure into a more competitive environment. Following this change in priorities, if regulation of a particular industry is still considered desirable, then greater attention should be paid to rate structures to ensure that output inefficiencies receive at least as much attention as the input distortions with which they are combined.

BIBLIOGRAPHY

Averch, H. and Johnson, L.L. "Behavior of the Firm Under Regulatory Constraint," *American Economic Review*, Vol. 52 (December 1962), pp. 1053-1069. ✓

Bailey, E.E. *Economic Theory of Regulatory Constraint*, Ph.D. Dissertation. Bell Telephone Laboratories and New York University, Revised June 1973.

______ and Coleman, R.D. "Discussion of Klevorick's Graduated Fair Return," Bell Laboratories Technical Memorandum, June 1970.

______ and Coleman, R.D. "The Effect of Lagged Regulation in an Averch-Johnson Model," *Bell Journal of Economics and Management Science*, Vol. 2 (Spring 1971), pp. 279-92.

______ and Malone, J.C. "Resource Allocation and the Regulated Firm," *Bell Journal of Economics and Management Science*, Vol. 1 (Spring 1970), pp. 129-42.

______ and White, L.J. "Reversals in Peak and Offpeak Prices," Manuscript (May 1973).

Baumol, W.J. *Business Behavior, Value and Growth*, rev. ed. New York: Harcourt, Brace & World, Inc., 1967.

✓ ______ and Klevorick, A.K. "Input Choices and Rate-of-Return Regulation: An Overview of the Discussion," *Bell Journal of Economics and Management Science*, Vol. 1 (Autumn 1970), pp. 162-90.

Clemens, E.W. "Price Discrimination and the Multiple-Product Firm," *The Review of Economic Studies*, Vol. 19 (1950-51), pp. 1-11.

Corey, G.R. "The Averch and Johnson Proposition: A Critical Analysis," *Bell Journal of Economics and Management Science*, Vol. 2 (Spring 1971), pp. 358-73.

Emery, E.D. "Regulated Utilities and Equipment Manufactures' Conspiracies in the Electric Power Industry," *Bell Journal of Economics and Management Science*, Vol. 4 (Spring 1973), pp. 322-37.

Hicks, J.R. Value and Capital, 2nd ed. Oxford: Clarendon Press, 1946.

Johnson, L.L. "Behavior of the Firm Under Regulatory Constraint: A Reassessment," American Economic Review, Vol. 63 (May 1973), pp. 90-97.

Kafoglis, M.Z. "Output of the Restrained Firm," American Economic Review, Vol. 59 (September 1969), pp. 583-89.

______ and Needy, C.W. "'Spread' in the Design of Rate Structures," Manuscript, 1972.

Kahn, A.E. "The Graduated Fair Return: Comment," American Economic Review, Vol. 58 (March 1968), pp. 170-73.

______. The Economics of Regulation, Vol. I and Vol. II, New York: John Wiley & Sons, 1970-71.

Kendrick, J.W. Productivity Trends in the United States. Princeton: Princeton University Press, 1961.

Klevorick, A.K. "The Graduated Fair Return: A Regulatory Proposal," American Economic Review, Vol. 56 (June 1968), pp. 477-84.

______. "The Graduated Fair Return: A Further Comment," American Economic Review, Vol. 61 (September 1971), pp. 727-29.

______. "The Behavior of a Firm Subject to Stochastic Regulatory Review," Bell Journal of Economics and Management Science, Vol. 4 (Spring 1973), pp. 57-88.

Knight, F. Risk, Uncertainty, and Profit. Reprints of Economic Classics. New York: Augustus M. Kelley, 1964.

Magee, S.P. "Factor Market Distortions, Production, and Trade: A Survey," Oxford Economic Papers, Vol. 25, No. 1 (March 1973), pp. 1-43.

McNicol, D.L. "The Comparative Statics Properties of the Theory of the Regulated Firm," Bell Journal of Economics and Management Science, Vol. 4, No. 2 (Autumn 1973), pp. 428-53.

Pace, J.D. "Relevant Markets and the Nature of Competition in the Electric Utility Industry," Antitrust Bulletin (Winter 1971), pp. 725-65.

Pigou, A.C. The Economics of Welfare. London: MacMillan and Company, 1960.

Reder, M.W. *Studies in the Theory of Welfare Economics*. New York: 1947.

Robinson, J. *Economics of Imperfect Competition*. London: MacMillan and Co., 1936.

Scherer, F.M. *Industrial Market Structure and Economic Performance*. Chicago: Rand McNally & Co., 1970.

Sheshinski, E. "Welfare Aspects of a Regulatory Constraint: Note," *American Economic Review*, Vol. 61 (March 1971), pp. 175-78.

Swan, P.L. "Market Structure and Technological Progress: The Influence of Monopoly on Product Innovation," *Quarterly Journal of Economics*, Vol. 84 (November 1970), pp. 627-38.

Wein, H.H. "Fair Rate of Return and Incentives--Some General Considerations," H.M. Trebing, ed. *Performance Under Regulation*. East Lansing: Michigan State University, 1968.

Wellisz, S.H. "Regulation of Natural Gas Pipeline Companies: An Economic Analysis," *Journal of Political Economy*, Vol. 71 (February 1963), pp. 30-43.

Westfield, F.M. "Regulation and Conspiracy," *American Economic Review*, Vol. 55 (June 1965), pp. 424-43.

White, L.J. "A Note on the Influence of Monopoly on Product Innovation," *Quarterly Journal of Economics*, Vol. 86 (May 1972), pp. 342-49.

Wichers, C.R. "The Graduated Fair Return: Comment," *American Economic Review*, Vol. 61 (September 1971), pp. 725-26.

Williamson, O.E. *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*. Englewood Cliffs: Prentice-Hall, 1964.

Zajac, E.E. "A Geometric Treatment of Averch-Johnson's Behavior of the Firm Model," *American Economic Review*, Vol. 60 (March 1970), pp. 117-25.

______. "Note on 'Gold Plating' or 'Rate Base Padding'," *Bell Journal of Economics and Management Science*, Vol. 3, No. 1 (Spring 1972), pp. 311-15.

BIOGRAPHICAL SKETCH

Charles William Needy

Personal Data: Born in Brawley, California, November 28, 1944, the son of Charles W. and Beulah L. Needy.

Education: Attended grade school in Brawley, California; was graduated from Smith-Cotton High School, Sedalia, Missouri, in 1962; received the Bachelor of Arts degree from the Southwest Missouri State College, with a major in Economics, in May 1966.

Received the Master of Science degree from the Oklahoma State University, with a major in Economics, in August 1967; completed requirements for the Doctor of Philosophy degree in June 1974.

Professional Experience: Was a surveyor and tabulator for a socioeconomic survey conducted in Springfield, Missouri, by the Sociology and Economics Department of Southwest Missouri State College during the fall semester, 1965.

Was a planning analyst for the Refined Products Planning and Corporate Planning Departments of Cities Service Oil Company in Tulsa during 1967 and 1968; was a supply analyst for the Crude Oil Purchases and Sales Department of Pan American Petroleum Corporation (with Standard Oil Company of Indiana) in Tulsa from 1968 to 1970.

Taught economics at Meramec Community College in St. Louis during the academic year 1970-71; served as a graduate teaching assistant in the Department of Economics at the University of Florida during 1972 and 1973; worked as a graduate research assistant in the Department of Economics at the University of Florida during the academic year, 1973-1974. Is a member of the American Economic Association and the Southern Economic Association.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Milton Z Kafoglis

Milton Z. Kafoglis, Chairman
Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sanford V Berg

Sanford V. Berg
Assistant Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Roger D Blair

Roger D. Blair
Assistant Professor of Economics

This dissertation was submitted to the Graduate Faculty of the Department of Economics in the College of Business Administration and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

June 1974

Dean, Graduate School